B2

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A01N 37/18, 43/04, A61K 38/00, C07H 21/02, C07K 1/00, 5/00, 16/00, C07G 17/00, C12N 1/20, 15/00, C12P 21/06, C12Q 1/68, G01N 33/53, H01R 13/62

A1

(11) International Publication Number: WO 97/27747

(43) International Publication Date: 7 August 1997 (07.08.97)

(21) International Application Number: PCT/US96/01640

(22) International Filing Date: 5 February 1996 (05.02.96)

(71) Applicant *(for all designated States except US)*: HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850-3338 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). GENTZ, Reiner, L. [DE/US]; 13404 Fairland Park Drive, Silver Spring, MD 20904 (US). YU, Guo-Liang [CN/US]; 13542 Straw Bale Lane, Darnestown, MD 20878 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). SU, Jeffrey [CA/US]; 443 Westside Drive #304, Gaithersburg, MD 20878 (US).

(74) Agent: BENSON, Robert, H.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).

(81) Designated States: AM, AU, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, JP, KG, KP, KR, KZ, LT, LV, MD, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, UA, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: CYTOSTATIN I

(57) Abstract

The invention relates to cytostatin I polypeptides, polynucleotides encoding the polypeptides, methods for producing the polypeptides, in particular by expressing the polynucleotides, and agonists and antagonists of the polypeptides. The invention further relates to methods for utilizing such polynucleotides, polypeptides, agonists and antagonists for applications, which relate, in part, to research, diagnostic and clinical arts.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

CYTOSTATIN I

This invention relates, in part, to newly identified polynucleotides and polypeptides; variants and derivatives of the polynucleotides and polypeptides; processes for making the polynucleotides and the polypeptides, and their variants and derivatives; agonists and antagonists of the polypeptides; and uses of the polynucleotides, polypeptides, variants, derivatives, agonists and antagonists. In particular, in these and in other regards, the invention relates to polynucleotides and polypeptides of human Cytostatin I.

BACKGROUND OF THE INVENTION

The cytostatin I of the present invention has been putatively identified as a growth inhibitory protein. This identification has been made as a result of amino acid sequence homology to mammary-derived growth inhibitor (MDGI) and direct measurements on cell growth.

Mammary-derived growth inhibitor (MDGI) is a cell growth inhibitor and differentiation factor firstly purified mammary carcinoma cells Ehrlich ascites, and then from cows milk and bovine mammary gland (Grosse et al. 2 references). MDGI inhibits proliferation of mammary epithelial cell lines in a dose-dependent and reversible manner. Maximal inhibition of cell proliferation by purified MDGI is in the range of 35 to 50%. In these cells

half-maximal inhibition was obtained with about 10^{-10} M MDGI (1 ng/ml). Inhibition was abolished by simultaneously adding epidermal growth factor (EGF), insulin. MDGI also inhibits the proliferation of several other permanent mammary carcinoma cell lines. MDGI has been shown to be immunologically related to a fibroblast growth inhibitor.

Peptides that locally signal growth cessation and stimulate differentiation of the developing epithelium are very important for mammary gland development. Recombinant and wild-type forms of mammary-derived growth inhibitor (MDGI) and heart-fatty acid binding protein (FABP), which belong to the FABP family, specifically inhibit growth of normal mouse mammary epithelial cells (MEC) and promote morphological differentiation, stimulates its own expression and promotes milk protein synthesis. Selective inhibition of endogenous MDGI expression in MEC by antisense phosphorothioate oligonucleotides suppresses appearance of alveolar end buds and lowers the beta-casein level in organ cultures. Furthermore, MDGI suppresses the mitogenic effects of EGF, and EGF antagonizes the activities of MDGI. Finally, the regulatory properties of MDGI can be fully mimicked by an 11-amino acid sequence, represented in the COOH terminus of MDGI and a subfamily of structurally related FABPs. MDGI is the first known growth inhibitor which promotes mammary gland differentiation. The amount of MDGI increased dramatically with the onset of lactation after delivery. Recent studies shows that a new posttranslational processing form of MDGI, MDGI 2, not present in lactation, was found in the bovine gland during pregnancy. (Brandt et al., Biochem. Biophy. Res. Comm., Vol. 189, p. 406, November 30, 1992.) To date, bovine, rat and mouse MDGI have been identified but no human MDGI or MDGI-like protein.

There is no sequence homology between MDGI and other known growth inhibitors. Thus, along with interferons, transforming growth factors β, and tumor necrosis factors, MDGI is one of the few naturally occurring growth inhibitors for mammary epithelium identifier so far.

Sequence analysis revealed extensive sequence homology of MDGI to a family of low molecular mass hydrophobic ligand-binding proteins, among them fatty acid-binding protein (FABP) from brain and heart, myelin P2, a differentiation associated protein in adipocytes (p422), gastrotropin, and the cellular retinoic acid-binding protein (CRABP). These proteins basically share two properties in common: they bind hydrophobic ligands such as long-chain fatty acids, retinoids, and eicosanoids, and they are expressed in a differentiation-dependent manner in mammary gland, heart, liver, brain, or intestine. All these proteins act intracellularly except MDGI and gastrotropin, which act extracellularly *in vitro*. The C-terminus of MDGI residues 126-130 are identical to residues 108-112 of bovine growth hormone. This stretch of amino acids is part of a sequence of growth hormone that is essential for its biological activity. Synthetic peptides corresponding to the MDGI-sequence, residue 121-131 mimic the effects of MDGI. The functions of these MDGI proteins are not yet well-defined, although a role in fatty acid transport, sequestration, or metabolism has been widely discussed. Interaction with as yet unknown hydrophobic ligands might play a functional role in the mechanism of growth inhibition exerted by MDGI. It is proposed that MDGI may act in an autocrine manner as a growth inhibitor, however, MDGI lacks a signal sequence for membrane translocation, most of MDGI has an intracellular localization. With regard to the secretion, an analogy might exist to other growth factors that also lack a signal sequence like FGF and PG-ECGF. In those cases cell damage as a possible way of secretion, or the existence of related factors with a signal sequence as a physiological ligands of the respective surface receptors, have been discussed.

Among other activities, MDGI reportedly may inhibit c-fos, c-myc and c-ras expression. MDGI has differentiation-promoting activity on mouse pluripotent embryonic stem cells and supports the commitment of undifferentiated ESC for neural differentiation. It is also suggested that MDGI

may be involved in the regulation of endothelial cell proliferation.

MDGI inhibits the induction of supersensitivity of neonatal rat heart muscle cells for beta-adrenergic receptors by lipoxygenase metabolites and various agents. The inhibitory activity of MDGI related to the induction of supersensitivity for hydrophilic beta-adrenergic agonists might point to a physiological role for a close relative of MDGI - the cardiac fatty acid-binding protein (H-FABP). One function of H-FABP could be to protect, the heart, under pathophysiological conditions, from lipoxygenase metabolites causing supersensitivity of beta-adrenergic receptors. Thus, H-FABP may be a physiological modulator of beta-adrenergic responses in the cardiac muscle. There is a need for a human MDGI-like protein and the gene encoding it.

SUMMARY OF THE INVENTION

Toward these ends, and others, it is an object of the present invention to provide polypeptides, inter alia, that have been identified as novel cytostatin I by homology between the amino acid sequence set out in Figure 1 and known amino acid sequences of other proteins such as mouse mammary-derived growth inhibitor (MDGI).

It is a further object of the invention, moreover, to provide polynucleotides that encode cytostatin I, particularly polynucleotides that encode the polypeptide herein designated cytostatin I.

In a particularly preferred embodiment of this aspect of the invention the polynucleotide comprises the region encoding human cytostatin I in the sequence set out in Figure 1.

In accordance with this aspect of the present invention there is provided an isolated nucleic acid molecule encoding a mature polypeptide expressed by the human cDNA contained in ATCC Deposit No. 97103.

In accordance with this aspect of the invention there are provided isolated nucleic acid molecules encoding human

cytostatin I, including mRNAs, cDNAs, genomic DNAs and, in further embodiments of this aspect of the invention, biologically, diagnostically, clinically or therapeutically useful variants, analogs or derivatives thereof, or fragments thereof, including fragments of the variants, analogs and derivatives.

Among the particularly preferred embodiments of this aspect of the invention are naturally occurring allelic variants of human cytostatin I.

It also is an object of the invention to provide cytostatin I polypeptides, particularly human cytostatin I polypeptides, that may be employed therapeutically as a cell growth inhibitor, to cause differentiation stimulatory activity on various responsive types of tissues and cells, to treat neoplasia, to inhibit angiogenesis, to inhibit metastases of tumor cells, to stimulate milk production and promote involution of the breast.

In accordance with this aspect of the invention there are provided novel polypeptides of human origin referred to herein as cytostatin I as well as biologically, diagnostically or therapeutically useful fragments, variants and derivatives thereof, variants and derivatives of the fragments, and analogs of the foregoing.

Among the particularly preferred embodiments of this aspect of the invention are variants of human cytostatin I encoded by naturally occurring alleles of the human cytostatin I gene.

It is another object of the invention to provide a process for producing the aforementioned polypeptides, polypeptide fragments, variants and derivatives, fragments of the variants and derivatives, and analogs of the foregoing. In a preferred embodiment of this aspect of the invention there are provided methods for producing the aforementioned cytostatin I polypeptides comprising culturing host cells having expressibly incorporated therein an exogenously-derived human cytostatin I-encoding polynucleotide under conditions for expression of human

cytostatin I in the host and then recovering the expressed polypeptide.

In accordance with another object the invention there are provided products, compositions, processes and methods that utilize the aforementioned polypeptides and polynucleotides for research, biological, clinical and therapeutic purposes, *inter alia*.

In accordance with certain preferred embodiments of this aspect of the invention, there are provided products, compositions and methods, *inter alia*, for, among other things: assessing cytostatin I expression in cells by determining cytostatin I polypeptides or cytostatin I-encoding mRNA; assaying genetic variation and aberrations, such as defects, in cytostatin I genes; and administering a cytostatin I polypeptide or polynucleotide to an organism to augment cytostatin I function or remediate cytostatin I dysfunction.

In accordance with certain preferred embodiments of this and other aspects of the invention there are provided probes that hybridize to human cytostatin I sequences.

In certain additional preferred embodiments of this aspect of the invention there are provided antibodies against cytostatin I polypeptides. In certain particularly preferred embodiments in this regard, the antibodies are highly selective for human cytostatin I.

In accordance with another aspect of the present invention, there are provided cytostatin I agonists. Among preferred agonists are molecules that mimic cytostatin I, that bind to cytostatin I-binding molecules or receptor molecules, and that elicit or augment cytostatin I-induced responses. Also among preferred agonists are molecules that interact with cytostatin I or cytostatin I polypeptides, or with other modulators of cytostatin I activities, and thereby potentiate or augment an effect of cytostatin I or more than one effect of cytostatin I.

In accordance with yet another aspect of the present invention, there are provided cytostatin I antagonists. Among preferred antagonists are those which mimic

cytostatin I so as to bind to cytostatin I receptor or binding molecules but not elicit a cytostatin I-induced response or more than one cytostatin I-induced response. Also among preferred antagonists are molecules that bind to or interact with cytostatin I so as to inhibit an effect of cytostatin I or more than one effect of cytostatin I or which prevent expression of cytostatin I.

In a further aspect of the invention there are provided compositions comprising a cytostatin I polynucleotide or a cytostatin I polypeptide for administration to cells in vitro, to cells ex vivo and to cells in vivo, or to a multicellular organism. In certain particularly preferred embodiments of this aspect of the invention, the compositions comprise a cytostatin I polynucleotide for expression of a cytostatin I polypeptide in a host organism for treatment of disease. Particularly preferred in this regard is expression in a human patient for treatment of a dysfunction associated with aberrant endogenous activity of cytostatin I.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill from the following description. It should be understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Nucleotide and deduced amino acid sequence of human Cytostatin I.

The nucleotide sequence of the cDNA encoding human cytostatin and amino acid sequence is shown. The cDNA sequence encodes a primary translation product of 107 amino acids of which the first 21 to 38 amino acids likely

represent a putative leader sequence or transmembrane domain.

Figure 2. Sequence homology of Cytostatin I with other family members (SEQ ID NO:7-11) .

Comparison of the amino acid sequence of cytostatin I (HTOBH93, top) to other members in the family is shown (SEQ ID NO:7-11).

Figure 3. Tissue distribution of cytostatin I.

(3A & 3B) Two μg of polyA RNA from the human tissues indicated were separated on a 1% agarose-formaldehyde gel and transferred to a nylon membrane. The membrane was probed with ^{32}P-labeled cytostatin I cDNA probe. Cytostatin I is highly expressed in spleen and kidney, moderately expressed in liver and thymus. The lanes on the 3A and 3B gels are:

Figure 3A	Figure 3B
Lane 1, spleen	heat
Lane 2, thymus	brain
Lane 3, prostate	placenta
Lane 4, testis	lung
Lane 5, ovary	liver
Lane 6, small intestine	skeletal muscle
Lane 7, colon	kidney
Lane 8, peripheral blood leukocytes	pancreas

RNA size market (kb): 9.5; 7.5; 4.4; 2.4; 1.35.

3C) 10 μg of total RNA from the cell lines shown were separated on a 1% agarose-formaldehyde gel and transferred to a nylon membrane. The membrane was probed with ^{32}P-labeled cytostatin I cDNA. Lane 1, CAMA1 (breast cancer); Lane 2 AN3CA (uterine cancer); Lane 3, SK.UT.1 (uterine cancer); Lane 4, MG63 (osteoblastoma); Lane 5, HOS (osteoblastoma; Lane 6, MCF7 (breast cancer); Lane 7, OVCAR-3 (ovarian cancer); Lane 8, CAOV-3 (ovarian cancer); Lane 9, HUVEC; Lane 10, AOSMIC (smooth muscle); Lane 11, Fore skin fibroblast. The expression of cystatin I is undetectable in these cells.

Figure 4. Purification of bacterial-expressed human cytostatin I (HG07400-2E).

The entire coding sequence including the putative signal sequence or transmembrane domain was fused in frame with a 6-His tag present in the expression vector pQE9 (Qiagen). *E. coli* haboring the expression plasmid were induced with 1 mM IPTG during the logarithmic growth phase. Following a 3-hour induction, the cell pellet was lysed with 6M Guanidine hydrochloride and cytostatin I was purified using a Nickel-chelate affinity chromatography column. The highly purified protein was denatured by dialysis in PBS buffer. M, molecular weight markers; Lane 1 and 2, induced cell lysate; Lane 3 and 4, uninduced cell lysate; Lane 5, pass through fraction from Nickel-chelate column purification; Lane 6, 7 and 8, Fraction eluted with 7M Guanidine hydrochloride (pH 5); 9 Fraction eluted with 6M Guanidine hydrochloride (pH 2).

Figure 5A Growth inhibitory activity of cytostatin I (HG07400-1E, highest concentration 100 ng/ml) against Mdamb 231 human breast cancer cells.

Figure 5B Growth inhibitory activity of cytostatin I (HG07400-2E, highest concentration 1000 ng/ml) against Mdamb 231 human breast cancer cells.

Figure 5C Growth inhibitory activity of cytostatin I (HG07400-1E) against Jurat human T cell leukemia cells.

Figure 5D Growth inhibitory activity of cytostatin I (HG07400-2E) against CCD-29LU human lung fibroblast cells.

Figure 5E Growth inhibitory activity of cytostatin I (HG07400-2E) against CPA 47 bovine pulmonary artery endothelial cells.

Figure 6. Northern blot analysis of cytostatin I expression in human breast tissues. Total RNAs were prepared from five metastatic breast carcinomas (C represents carcinomas) and five benign breasts (B represents benign breast). RNA samples from B1-B4 were isolated from breast fibroadenomas, and RNA sample of B5 was isolated from breast hyperplasia. Each lane contained 30 ug of total RNA. (A) TMP-4 RNA hybridized with

^{32}P-labeled full-length cytostatin I cDNA probe. (B) 18 S rRNA indicating the integrity of the RNA samples and the loading control.

Figure 7. *In situ* hybridization analysis of cytostatin I (A-D) expression in human breast. Open arrows indicate the stromal cells and closed arrows indicate both normal and neoplastic breast epithelial cells. Areas with brown color indicate the cytostatin I signals. **(A)** Low magnification (100X) view of fibroadenomas showing a strong labeling of epithelial cells for cytostatin I mRNA. **(B)** Low-power view (100X) of hyperplasia shows a negative staining for cytostatin I. (C) Negatively stained low grade *in situ* carcinoma. **(D)** Infiltrating carcinoma- low magnification view (160X) of the negatively stained cytostatin I . All the Sections were counterstained lightly with hematoxylin for a better view of the negatively stained of malignant and highly proliferative breast epithelial cells.

GLOSSARY

The following illustrative explanations are provided to facilitate understanding of certain terms used frequently herein, particularly in the examples. The explanations are provided as a convenience and are not limitative of the invention.

DIGESTION of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan.

For analytical purposes, typically, 1 μg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 μl of reaction buffer. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes.

Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers.

Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well known methods that are routine for those skilled in the art.

GENETIC ELEMENT generally means a polynucleotide comprising a region that encodes a polypeptide or a region that regulates transcription or translation or other processes important to expression of the polypeptide in a hot cell, or a polynucleotide comprising both a region that encodes a polypeptide and a region operably linked thereto that regulates expression.

Genetic elements may be comprised within a vector that replicates as an episomal element; that is, as a molecule physically independent of the host cell genome. They may be comprised within mini-chromosomes, such as those that arise during amplification of transfected DNA by methotrexate selection in eukaryotic cells. Genetic elements also may be comprised within a host cell genome; not in their natural state but, rather, following manipulation such as isolation, cloning and introduction into a host cell in the form of purified DNA or in a vector, among others.

ISOLATED means altered "by the hand of man" from its natural state; i.e., that, if it occurs in nature, it has been changed or removed from its original environment, or both.

For example, a naturally occurring polynucleotide or a polypeptide naturally present in a living animal in its natural state is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting

materials of its natural state is "isolated", as the term is employed herein. For example, with respect to polynucleotides, the term isolated means that it is separated from the chromosome and cell in which it naturally occurs.

As part of or following isolation, such polynucleotides can be joined to other polynucleotides, such as DNAs, for mutagenesis, to form fusion proteins, and for propagation or expression in a host, for instance. The isolated polynucleotides, alone or joined to other polynucleotides such as vectors, can be introduced into host cells, in culture or in whole organisms. Introduced into host cells in culture or in whole organisms, such DNAs still would be isolated, as the term is used herein, because they would not be in their naturally occurring form or environment. Similarly, the polynucleotides and polypeptides may occur in a composition, such as a media formulations, solutions for introduction of polynucleotides or polypeptides, for example, into cells, compositions or solutions for chemical or enzymatic reactions, for instance, which are not naturally occurring compositions, and, therein remain isolated polynucleotides or polypeptides within the meaning of that term as it is employed herein.

LIGATION refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double stranded DNAs. Techniques for ligation are well known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, for instance, Sambrook et al., MOLECULAR CLONING, A LABORATORY MANUAL, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989) and Maniatis et al., pg. 146, as cited below.

OLIGONUCLEOTIDE(S) refers to relatively short polynucleotides. Often the term refers to single-stranded deoxyribonucleotides, but it can refer as well to single-or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including in vitro recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms.

Initially, chemically synthesized DNAs typically are obtained without a 5′ phosphate. The 5′ ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP.

The 3′ end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5′ phosphate of another polynucleotide, such as another oligonucleotide. As is well known, this reaction can be prevented selectively, where desired, by removing the 5′ phosphates of the other polynucleotide(s) prior to ligation.

PLASMIDS generally are designated herein by a lower case p preceded and/or followed by capital letters and/or numbers, in accordance with standard naming conventions that are familiar to those of skill in the art.
Starting plasmids disclosed herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids by routine application of well known, published procedures. Many plasmids and other cloning and expression vectors that can be used in accordance with the present invention are well known and readily available to those of skill in the art. Moreover, those of skill readily may construct any number of other plasmids suitable for use in the invention. The properties, construction and use of such plasmids, as well

as other vectors, in the present invention will be readily apparent to those of skill from the present disclosure.

POLYNUCLEOTIDE(S) generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. Thus, for instance, polynucleotides as used herein refers to, among others, single-and double-stranded DNA, DNA that is a mixture of single-and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, polynucleotide as used herein refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The strands in such regions may be from the same molecule or from different molecules. The regions may include all of one or more of the molecules, but more typically involve only a region of some of the molecules. One of the molecules of a triple-helical region often is an oligonucleotide.

As used herein, the term polynucleotide includes DNAs or RNAs as described above that contain one or more modified bases. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein.

It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including simple and complex cells, inter alia.

POLYPEPTIDES, as used herein, includes all polypeptides as described below. The basic structure of

polypeptides is well known and has been described in innumerable textbooks and other publications in the art. In this context, the term is used herein to refer to any peptide or protein comprising two or more amino acids joined to each other in a linear chain by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types.

It will be appreciated that polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids, and that many amino acids, including the terminal amino acids, may be modified in a given polypeptide, either by natural processes, such as processing and other post-translational modifications, but also by chemical modification techniques which are well known to the art. Even the common modifications that occur naturally in polypeptides are too numerous to list exhaustively here, but they are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature, and they are well known to those of skill in the art.

Among the known modifications which may be present in polypeptides of the present are, to name an illustrative few, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation,

sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York (1983); Seifter et al., Analysis for protein modifications and nonprotein cofactors, Meth. Enzymol. 182: 626-646 (1990) and Rattan et al., Protein Synthesis: Posttranslational Modifications and Aging, Ann. N.Y. Acad. Sci. 663: 48-62 (1992).

It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well.

Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well.

For instance, the amino terminal residue of polypeptides made in E. coli, prior to proteolytic processing, almost invariably will be N-formylmethionine.

The modifications that occur in a polypeptide often will be a function of how it is made. For polypeptides made by expressing a cloned gene in a host, for instance, the nature and extent of the modifications in large part will be determined by the host cell posttranslational modification capacity and the modification signals present in the polypeptide amino acid sequence. For instance, as is well known, glycosylation often does not occur in bacterial hosts such as E. coli. Accordingly, when glycosylation is desired, a polypeptide should be expressed in a glycosylating host, generally a eukaryotic cell. Insect cell often carry out the same posttranslational glycosylations as mammalian cells and, for this reason, insect cell expression systems have been developed to express efficiently mammalian proteins having native patterns of glycosylation, inter alia. Similar considerations apply to other modifications.

It will be appreciated that the same type of modification may be present in the same or varying degree at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications.

In general, as used herein, the term polypeptide encompasses all such modifications, particularly those that are present in polypeptides synthesized by expressing a polynucleotide in a host cell.

VARIANT(S) of polynucleotides or polypeptides, as the term is used herein, are polynucleotides or polypeptides that differ from a reference polynucleotide or polypeptide, respectively. Variants in this sense are described below and elsewhere in the present disclosure in greater detail.

(1) A polynucleotide that differs in nucleotide sequence from another, reference polynucleotide. Generally, differences are limited so that the nucleotide

sequences of the reference and the variant are closely similar overall and, in many regions, identical.

As noted below, changes in the nucleotide sequence of the variant may be silent. That is, they may not alter the amino acids encoded by the polynucleotide. Where alterations are limited to silent changes of this type a variant will encode a polypeptide with the same amino acid sequence as the reference. Also as noted below, changes in the nucleotide sequence of the variant may alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Such nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below.

(2) A polypeptide that differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference and the variant are closely similar overall and, in many region, identical.

A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions, fusions and truncations, which may be present in any combination.

RECEPTOR MOLECULE, as used herein, refers to molecules which bind or interact specifically with cytostatin I polypeptides of the present invention, including not only classic receptors, which are preferred, but also other molecules that specifically bind to or interact with polypeptides of the invention (which also may be referred to as "binding molecules" and "interaction molecules," respectively and as "cytostatin I binding molecules" and "cytostatin I interaction molecules." Binding between polypeptides of the invention and such molecules, including receptor or binding or interaction molecules may be exclusive to polypeptides of the invention, which is very highly preferred, or it may be highly specific for polypeptides of the invention, which is highly preferred, or it may be highly specific to a group of proteins that

includes polypeptides of the invention, which is preferred, or it may be specific to several groups of proteins at least one of which includes polypeptides of the invention.

Receptors also may be non-naturally occurring, such as antibodies and antibody-derived reagents that bind specifically to polypeptides of the invention.

DESCRIPTION OF THE INVENTION

The present invention relates to novel cytostatin I polypeptides and polynucleotides, among other things, as described in greater detail below. In particular, the invention relates to polypeptides and polynucleotides of a novel human cytostatin I, which is related by amino acid sequence homology to mouse mammary-derived growth inhibitor. The invention relates especially to cytostatin I having the nucleotide and amino acid sequences set out in Figure 1, and to the cytostatin I nucleotide and amino acid sequences of the human cDNA in ATCC Deposit No. 97103, which is herein referred to as "the deposited clone" or as the "cDNA of the deposited clone." It will be appreciated that the nucleotide and amino acid sequences set out in Figure 1 were obtained by sequencing the cDNA of the deposited clone. Hence, the sequence of the deposited clone is controlling as to any discrepancies between the two and any reference to the sequences of Figure 1 include reference to the sequence of the human cDNA of the deposited clone.

Polynucleotides

In accordance with one aspect of the present invention, there are provided isolated polynucleotides which encode the cytostatin I polypeptide having the deduced amino acid sequence of Figure 1.

Using the information provided herein, such as the polynucleotide sequence set out in Figure 1, a polynucleotide of the present invention encoding human cytostatin I polypeptide may be obtained using standard cloning and screening procedures, such as those for cloning

cDNAs using mRNA from cells of human tissue as starting material. Human cytostatin I of the invention is structurally related to other proteins of the cytostatin family, as shown by the results of sequencing the cDNA encoding human cytostatin I in the deposited clone. The cDNA sequence thus obtained is set out in Figure 1.

MDGI was originally identified as the cellular retinoic acid-binding protein (CRABP). Both CRABP and MDGI belong to a family of proteins known to bind hydrophobic ligands, referred to as Fatty acid binding proteins (FABPs). Cytostatin I is 33% identical and 63% similar to mouse MDGI. Cytostatin I is highly expressed in spleen and kidney, moderately expressed in liver and thymus. The selective expression of cytostatin I was demonstrated during analysis expression in selected human tissues. The cytostatin I gene was found three times in nine week old early state library, it was found once each in breast lympho node library, pancreas library and tonsils library. Cytostatin I protein was expressed and purified from *E. coli*. Our findings demonstrate that cytostatin I has growth inhibitory activity against breast cancer cells, leukemia cells, fibroblast cells, and endothelial cells.

The coding sequence which encodes the polypeptide may be identical to the coding sequence of the polynucleotide shown in Figure 1. It also may be a polynucleotide with a different sequence, which, as a result of the redundancy (degeneracy) of the genetic code, encodes the polypeptide of the DNA of Figure 1.

Polynucleotides of the present invention which encode the polypeptide of Figure 1 may include, but are not limited to the coding sequence for the mature polypeptide, by itself; the coding sequence for the mature polypeptide and additional coding sequences, such as those encoding a leader or secretory sequence, such as a pre-, or pro- or prepro- protein sequence; the coding sequence of the mature polypeptide, with or without the aforementioned additional coding sequences, together with additional, non-coding sequences, including for example, but not limited to

introns and non-coding 5' and 3' sequences, such as the transcribed, non-translated sequences that play a role in transcription, mRNA processing - including splicing and polyadenylation signals, for example - ribosome binding and stability of mRNA; additional coding sequence which codes for additional amino acids, such as those which provide additional functionalities. Thus, for instance, the polypeptide may be fused to a marker sequence, such as a peptide, which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, such as the tag provided in the vector pQE-9, among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci., USA 86: 821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein.

The HA tag corresponds to an epitope derived of influenza hemagglutinin protein, which has been described by Wilson et al., Cell 37: 767 (1984), for instance.

In accordance with the foregoing, the term "polynucleotide encoding a polypeptide" as used herein encompasses polynucleotides which include a sequence encoding a polypeptide of the present invention, particularly the human cytostatin I having the amino acid sequence set out in Figure 1. The term encompasses polynucleotides that include a single continuous region or discontinuous regions encoding the polypeptide (for example, interrupted by introns) together with additional regions, that also may contain coding and/or non-coding sequences.

The present invention further relates to variants of the herein above described polynucleotides which encode for fragments, analogs and derivatives of the polypeptide having the deduced amino acid sequence of Figure 1. A variant of the polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide

may be made by mutagenesis techniques, including those applied to polynucleotides, cells or organisms.

Among variants in this regard are variants that differ from the aforementioned polynucleotides by nucleotide substitutions, deletions or additions. The substitutions, deletions or additions may involve one or more nucleotides. The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

Among the particularly preferred embodiments of the invention in this regard are polynucleotides encoding polypeptides having the amino acid sequence of cytostatin I set out in Figure 1 or the amino acid sequence of cytostatin I of the cDNA of the deposited clone; variants, analogs, derivatives and fragments thereof, and fragments of the variants, analogs and derivatives.

Further particularly preferred in this regard are polynucleotides encoding cytostatin I variants, analogs, derivatives and fragments, and variants, analogs and derivatives of the fragments, which have the amino acid sequence of the cytostatin I polypeptide of Figure 1 in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the cytostatin I. Also especially preferred in this regard are conservative substitutions. Most highly preferred are polynucleotides encoding polypeptides having the amino acid sequence of Figure 1 without substitutions.

Further preferred embodiments of the invention are polynucleotides that are at least 70% identical to a polynucleotide encoding the cytostatin I polypeptide having the amino acid sequence set out in Figure 1, and polynucleotides which are complementary to such polynucleotides. Alternatively, most highly preferred are polynucleotides that comprise a region that is at least 80%

identical to a polynucleotide encoding the cytostatin I polypeptide of the cDNA of the deposited clone and polynucleotides complementary thereto. In this regard, polynucleotides at least 90% identical to the same are particularly preferred, and among these particularly preferred polynucleotides, those with at least 95% are especially preferred. Furthermore, those with at least 97% are highly preferred among those with at least 95%, and among these those with at least 98% and at least 99% are particularly highly preferred, with at least 99% being the more preferred.

Particularly preferred embodiments in this respect, moreover, are polynucleotides which encode polypeptides which retain substantially the same biological function or activity as the mature polypeptide encoded by the cDNA of Figure 1.

The present invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences.

As discussed additionally herein regarding polynucleotide assays of the invention, for instance, polynucleotides of the invention as discussed above, may be used as a hybridization probe for cDNA and genomic DNA to isolate full-length cDNAs and genomic clones encoding cytostatin I and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the human cytostatin I gene. Such probes generally will comprise at least 15 bases. Preferably, such probes will have at least 30 bases and may have at least 50 bases. Particularly preferred probes will have at least 30 bases and will have 50 bases or less.

For example, the coding region of the cytostatin I gene may be isolated by screening using the known DNA sequence to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the present invention is then used to screen a library of human cDNA, genomic DNA or mRNA to determine which members of the library the probe hybridizes to.

The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of treatments and diagnostics to human disease, as further discussed herein relating to polynucleotide assays, inter alia.

The polynucleotides may encode a polypeptide which is the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature polypeptide (when the mature form has more than one polypeptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, may facilitate protein trafficking, may prolong or shorten protein half-life or may facilitate manipulation of a protein for assay or production, among other things. As generally is the case in situ, the additional amino acids may be processed away from the mature protein by cellular enzymes.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. When prosequences are removed such inactive precursors generally are activated. Some or all of the prosequences may be removed before activation. Generally, such precursors are called proproteins.

Deposited materials

A deposit containing a human cytostatin I cDNA has been deposited with the American Type Culture Collection, as noted above. Also as noted above, the cDNA deposit is referred to herein as "the deposited clone" or as "the cDNA of the deposited clone."

The deposited clone was deposited with the American Type Culture Collection, 12301 Park Lawn Drive, Rockville, Maryland 20852, USA, on March 21, 1995, and assigned ATCC Deposit No. 97103.

The deposited material is a pBluescript SK (-) plasmid (Stratagene, La Jolla, CA) that contains the full length cytostatin I cDNA.

The deposit has been made under the terms of the Budapest Treaty on the international recognition of the deposit of micro-organisms for purposes of patent procedure. The strain will be irrevocably and without restriction or condition released to the public upon the issuance of a patent. The deposit is provided merely as convenience to those of skill in the art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. §112.

The sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

A license may be required to make, use or sell the deposited materials, and no such license is hereby granted.

Polypeptides

The present invention further relates to a human cytostatin I polypeptide which has the deduced amino acid sequence of Figure 1.

The invention also relates to fragments, analogs and derivatives of these polypeptides. The terms "fragment," "derivative" and "analog" when referring to the polypeptide of Figure 1 means a polypeptide which retains essentially the same biological function or activity as such polypeptide. Thus, an analog includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a

synthetic polypeptide. In certain preferred embodiments it is a recombinant polypeptide.

The fragment, derivative or analog of the polypeptide of Figure 1 may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the mature polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the mature polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the mature polypeptide or a proprotein sequence. Such fragments, derivatives and analogs are deemed to be within the scope of those skilled in the art from the teachings herein.

Among the particularly preferred embodiments of the invention in this regard are polypeptides having the amino acid sequence of cytostatin I set out in Figure 1, variants, analogs, derivatives and fragments thereof, and variants, analogs and derivatives of the fragments. Alternatively, particularly preferred embodiments of the invention in this regard are polypeptides having the amino acid sequence of the cytostatin I of the cDNA in the deposited clone, variants, analogs, derivatives and fragments thereof, and variants, analogs and derivatives of the fragments.

Among preferred variants are those that vary from a reference by conservative amino acid substitutions. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange

of the acidic residues Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe, Tyr.

Further particularly preferred in this regard are variants, analogs, derivatives and fragments, and variants, analogs and derivatives of the fragments, having the amino acid sequence of the cytostatin I polypeptide of Figure 1 in which several, a few, 5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues are substituted, deleted or added, in any combination. Especially preferred among these are silent substitutions, additions and deletions, which do not alter the properties and activities of the cytostatin I. Also especially preferred in this regard are conservative substitutions. Most highly preferred are polypeptides having the amino acid sequence of Figure 1 without substitutions.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The polypeptides of the present invention include the polypeptide of SEQ ID NO:2 (in particular the mature polypeptide) as well as polypeptides which have at least 70% similarity (preferably at least 70% identity) to the polypeptide of SEQ ID NO:2 and more preferably at least 90% similarity (more preferably at least 90% identity) to the polypeptide of SEQ ID NO:2 and still more preferably at least 95% similarity (still more preferably at least 95% identity) to the polypeptide of SEQ ID NO:2 and also include portions of such polypeptides with such portion of the polypeptide generally containing at least 30 amino acids and more preferably at least 50 amino acids.

As known in the art "similarity" between two polypeptides is determined by comparing the amino acid sequence and its conserved amino acid substitutes of one polypeptide to the sequence of a second polypeptide.

Fragments or portions of the polypeptides of the present invention may be employed for producing the

corresponding full-length polypeptide by peptide synthesis; therefore, the fragments may be employed as intermediates for producing the full-length polypeptides. Fragments or portions of the polynucleotides of the present invention may be used to synthesize full-length polynucleotides of the present invention.

Fragments

Also among preferred embodiments of this aspect of the present invention are polypeptides comprising fragments of cytostatin I, most particularly fragments of the cytostatin I having the amino acid set out in Figure 1, or having the amino acid sequence of the cytostatin I of the deposited clone, and fragments of variants and derivatives of the cytostatin I of Figure 1.

In this regard a fragment is a polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned cytostatin I polypeptides and variants or derivatives thereof.

Such fragments may be "free-standing," i.e., not part of or fused to other amino acids or polypeptides, or they may be comprised within a larger polypeptide of which they form a part or region. When comprised within a larger polypeptide, the presently discussed fragments most preferably form a single continuous region. However, several fragments may be comprised within a single larger polypeptide. For instance, certain preferred embodiments relate to a fragment of a cytostatin I polypeptide of the present comprised within a precursor polypeptide designed for expression in a host and having heterologous pre and pro-polypeptide regions fused to the amino terminus of the cytostatin I fragment and an additional region fused to the carboxyl terminus of the fragment. Therefore, fragments in one aspect of the meaning intended herein, refers to the portion or portions of a fusion polypeptide or fusion protein derived from cytostatin I.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which have from about 25 to about 107 amino acids.

In this context about includes the particularly recited range and ranges larger or smaller by several, a few, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes. For instance, about 25-107 amino acids in this context means a polypeptide fragment of 25 plus or minus several, a few, 5, 4, 3, 2 or 1 amino acids to 107 plus or minus several a few, 5, 4, 3, 2 or 1 amino acid residues, i.e., ranges as broad as 25 minus several amino acids to 107 plus several amino acids to as narrow as 25 plus several amino acids to 107 minus several amino acids.

Highly preferred in this regard are the recited ranges plus or minus as many as 5 amino acids at either or at both extremes. Particularly highly preferred are the recited ranges plus or minus as many as 3 amino acids at either or at both the recited extremes. Especially particularly highly preferred are ranges plus or minus 1 amino acid at either or at both extremes or the recited ranges with no additions or deletions. Most highly preferred of all in this regard are fragments from about 25 to about 107.

Among especially preferred fragments of the invention are truncation mutants of cytostatin I. Truncation mutants include cytostatin I polypeptides having the amino acid sequence of Figure 1, or of variants or derivatives thereof, except for deletion of a continuous series of residues (that is, a continuous region, part or portion) that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or, as in double truncation mutants, deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Fragments having the size ranges set out about also are preferred embodiments of truncation fragments, which are especially preferred among fragments generally.

Also preferred in this aspect of the invention are fragments characterized by structural or functional

attributes of cytostatin I. Preferred embodiments of the invention in this regard include fragments that comprise alpha-helix and alpha-helix forming regions ("alpha-regions"), beta-sheet and beta-sheet-forming regions ("beta-regions"), turn and turn-forming regions ("turn-regions"), coil and coil-forming regions ("coil-regions"), hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions and high antigenic index regions of cytostatin I.

Certain preferred regions in these regards are set out in Figure 3, and include, but are not limited to, regions of the aforementioned types identified by analysis of the amino acid sequene set out in Figure 1. As set out in Figure 3, such preferred regions include Garnier-Robson alpha-regions, beta-regions, turn-regions and coil-regions, Chou-Fasman alpha-regions, beta-regions and turn-regions, Kyte-Doolittle hydrophilic regions and hydrophilic regions, Eisenberg alpha and beta amphipathic regions, Karplus-Schulz flexible regions, Emini surface-forming regions and Jameson-Wolf high antigenic index regions.

Among highly preferred fragments in this regard are those that comprise regions of cytostatin I that combine several structural features, such as several of the features set out above. In this regard, the regions defined by the residues about 25 to about 107 of Figure 1, which all are characterized by amino acid compositions highly characteristic of turn-regions, hydrophilic regions, flexible-regions, surface-forming regions, and high antigenic index-regions, are especially highly preferred regions. Such regions may be comprised within a larger polypeptide or may be by themselves a preferred fragment of the present invention, as discussed above. It will be appreciated that the term "about" as used in this paragraph has the meaning set out above regarding fragments in general.

Further preferred regions are those that mediate activities of cytostatin I. Most highly preferred in this

regard are fragments that have a chemical, biological or other activity of cytostatin I, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Highly preferred in this regard are fragments that contain regions that are homologs in sequence, or in position, or in both sequence and to active regions of related polypeptides, such as the related polypeptides set out in Figure 2, which includes cytostatins. Among particularly preferred fragments in these regards are truncation mutants, as discussed above.

It will be appreciated that the invention also relates to, among others, polynucleotides encoding the aforementioned fragments, polynucleotides that hybridize to polynucleotides encoding the fragments, particularly those that hybridize under stringent conditions, and polynucleotides, such as PCR primers, for amplifying polynucleotides that encode the fragments. In these regards, preferred polynucleotides are those that correspondent to the preferred fragments, as discussed above.

Vectors, host cells, expression

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

Host cells can be genetically engineered to incorporate polynucleotides and express polypeptides of the present invention. For instance, polynucleotides may be introduced into host cells using well known techniques of infection, transduction, transfection, transvection and transformation. The polynucleotides may be introduced alone or with other polynucleotides. Such other polynucleotides may be introduced independently, co-introduced or introduced joined to the polynucleotides of the invention.

Thus, for instance, polynucleotides of the invention may be transfected into host cells with another, separate, polynucleotide encoding a selectable marker, using standard techniques for co-transfection and selection in, for instance, mammalian cells. In this case the polynucleotides generally will be stably incorporated into the host cell genome.

Alternatively, the polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. The vector construct may be introduced into host cells by the aforementioned techniques. Generally, a plasmid vector is introduced as DNA in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. Electroporation also may be used to introduce polynucleotides into a host. If the vector is a virus, it may be packaged in vitro or introduced into a packaging cell and the packaged virus may be transduced into cells. A wide variety of techniques suitable for making polynucleotides and for introducing polynucleotides into cells in accordance with this aspect of the invention are well known and routine to those of skill in the art. Such techniques are reviewed at length in Sambrook et al. cited above, which is illustrative of the many laboratory manuals that detail these techniques. In accordance with this aspect of the invention the vector may be, for example, a plasmid vector, a single or double-stranded phage vector, a single or double-stranded RNA or DNA viral vector. Such vectors may be introduced into cells as polynucleotides, preferably DNA, by well known techniques for introducing DNA and RNA into cells. The vectors, in the case of phage and viral vectors also may be and preferably are introduced into cells as packaged or encapsidated virus by well known techniques for infection and transduction. Viral vectors may be replication competent or replication defective. In the latter case viral propagation generally will occur only in complementing host cells.

Preferred among vectors, in certain respects, are those for expression of polynucleotides and polypeptides of the present invention. Generally, such vectors comprise cis-acting control regions effective for expression in a host operatively linked to the polynucleotide to be expressed. Appropriate trans-acting factors either are supplied by the host, supplied by a complementing vector or supplied by the vector itself upon introduction into the host.

In certain preferred embodiments in this regard, the vectors provide for specific expression. Such specific expression may be inducible expression or expression only in certain types of cells or both inducible and cell-specific. Particularly preferred among inducible vectors are vectors that can be induced for expression by environmental factors that are easy to manipulate, such as temperature and nutrient additives. A variety of vectors suitable to this aspect of the invention, including constitutive and inducible expression vectors for use in prokaryotic and eukaryotic hosts, are well known and employed routinely by those of skill in the art.

The engineered host cells can be cultured in conventional nutrient media, which may be modified as appropriate for, inter alia, activating promoters, selecting transformants or amplifying genes. Culture conditions, such as temperature, pH and the like, previously used with the host cell selected for expression generally will be suitable for expression of polypeptides of the present invention as will be apparent to those of skill in the art.

A great variety of expression vectors can be used to express a polypeptide of the invention. Such vectors include chromosomal, episomal and virus-derived vectors e.g., vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and

vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids, all may be used for expression in accordance with this aspect of the present invention. Generally, any vector suitable to maintain, propagate or express polynucleotides to express a polypeptide in a host may be used for expression in this regard.

The appropriate DNA sequence may be inserted into the vector by any of a variety of well-known and routine techniques. In general, a DNA sequence for expression is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction endonucleases and then joining the restriction fragments together using T4 DNA ligase. Procedures for restriction and ligation that can be used to this end are well known and routine to those of skill. Suitable procedures in this regard, and for constructing expression vectors using alternative techniques, which also are well known and routine to those skill, are set forth in great detail in Sambrook et al. cited elsewhere herein.

The DNA sequence in the expression vector is operatively linked to appropriate expression control sequence(s), including, for instance, a promoter to direct mRNA transcription. Representatives of such promoters include the phage lambda PL promoter, the E. coli lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name just a few of the well-known promoters. It will be understood that numerous promoters not mentioned are suitable for use in this aspect of the invention are well known and readily may be employed by those of skill in the manner illustrated by the discussion and the examples herein.

In general, expression constructs will contain sites for transcription initiation and termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the constructs will include a translation

initiating AUG at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

In addition, the constructs may contain control regions that regulate as well as engender expression. Generally, in accordance with many commonly practiced procedures, such regions will operate by controlling transcription, such as repressor binding sites and enhancers, among others.

The vector containing the appropriate DNA sequence as described elsewhere herein, as well as an appropriate promoter, and other appropriate control sequences, may be introduced into an appropriate host using a variety of well known techniques suitable to expression therein of a desired polypeptide. Representative examples of appropriate hosts include bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS and Bowes melanoma cells; and plant cells. Hosts for of a great variety of expression constructs are well known, and those of skill will be enabled by the present disclosure readily to select a host for expressing a polypeptides in accordance with this aspect of the present invention.

The following vectors, which are commercially available, are provided by way of example. Among vectors preferred for use in bacteria are pQE70, pQE60 and pQE-9, available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. These vectors are listed solely by way of illustration of the many commercially available and well known vectors that are available to those of skill in the art for use in accordance with this aspect of the present invention. It will be appreciated

that any other plasmid or vector suitable for, for example, introduction, maintenance, propagation or expression of a polynucleotide or polypeptide of the invention in a host may be used in this aspect of the invention.

Promoter regions can be selected from any desired gene using vectors that contain a reporter transcription unit lacking a promoter region, such as a chloramphenicol acetyl transferase ("cat") transcription unit, downstream of restriction site or sites for introducing a candidate promoter fragment; i.e., a fragment that may contain a promoter. As is well known, introduction into the vector of a promoter-containing fragment at the restriction site upstream of the cat gene engenders production of CAT activity, which can be detected by standard CAT assays. Vectors suitable to this end are well known and readily available. Two such vectors are pKK232-8 and pCM7. Thus, promoters for expression of polynucleotides of the present invention include not only well known and readily available promoters, but also promoters that readily may be obtained by the foregoing technique, using a reporter gene.

Among known bacterial promoters suitable for expression of polynucleotides and polypeptides in accordance with the present invention are the E. coli lacI and lacZ and promoters, the T3 and T7 promoters, the T5 tac promoter, the lambda PR, PL promoters and the trp promoter. Among known eukaryotic promoters suitable in this regard are the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Selection of appropriate vectors and promoters for expression in a host cell is a well known procedure and the requisite techniques for expression vector construction, introduction of the vector into the host and expression in the host are routine skills in the art.

The present invention also relates to host cells containing the above-described constructs discussed above.

The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al. BASIC METHODS IN MOLECULAR BIOLOGY, (1986).

Constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Mature proteins can be expressed in mammalian cells, yeast, bacteria, or other cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989).

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Polynucleotides of the invention, encoding the heterologous structural sequence of a polypeptide of the invention generally will be inserted into the vector using standard techniques so that it is operably linked to the promoter for expression. The polynucleotide will be positioned so that the transcription start site is located appropriately 5' to a ribosome binding site. The ribosome binding site will be 5' to the AUG that initiates translation of the polypeptide to be expressed. Generally, there will be no other open reading frames that begin with an initiation codon, usually AUG, and lie between the ribosome binding site and the initiating AUG. Also, generally, there will be a translation stop codon at the end of the polypeptide and there will be a polyadenylation signal and a transcription termination signal appropriately disposed at the 3' end of the transcribed region.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signals may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, region also may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to polypeptides to engender secretion or excretion, to improve stability and to facilitate purification, among others, are familiar and routine techniques in the art.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, where the selected promoter is inducible it is induced by appropriate means (e.g., temperature shift or exposure to chemical inducer) and cells are cultured for an additional period.

Cells typically then are harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well know to those skilled in the art.

Various mammalian cell culture systems can be employed for expression, as well. Examples of mammalian expression systems include the COS-7 lines of monkey kidney fibroblast, described in Gluzman et al., Cell 23: 175 (1981). Other cell lines capable of expressing a compatible vector include for example, the C127, 3T3, CHO, HeLa, human kidney 293 and BHK cell lines.

The cytostatin I polypeptide can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification. Well known techniques for refolding protein may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and

mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

Cytostatin I polynucleotides and polypeptides may be used in accordance with the present invention for a variety of applications, particularly those that make use of the chemical and biological properties cytostatin I. Among these are applications in tumor treatment, inhibition of angiogenesis, inhibition of metastases, stimulation of milk production and promotion of involution of the breast. Additional applications relate to diagnosis and to treatment of disorders of cells, tissues and organisms. These aspects of the invention are illustrated further by the following discussion.

Polynucleotide assays

This invention is also related to the use of the cytostatin I polynucleotides to detect complementary polynucleotides such as, for example, as a diagnostic reagent. Detection of a mutated form of cytostatin I associated with a dysfunction will provide a diagnostic tool that can add or define a diagnosis of a disease or susceptibility to a disease which results from under-expression over-expression or altered expression of cytostatin I, such as, for example, aberrant cellular proliferation.

Individuals carrying mutations in the human cytostatin I gene may be detected at the DNA level by a variety of techniques. Nucleic acids for diagnosis may be obtained from a patient's cells, such as from blood, urine, saliva, tissue biopsy and autopsy material. The genomic DNA may be used directly for detection or may be amplified enzymatically by using PCR prior to analysis. PCR (Saiki et al., Nature, 324: 163-166 (1986)). RNA or cDNA may also be used in the same ways. As an example, PCR primers

complementary to the nucleic acid encoding cytostatin I can be used to identify and analyze cytostatin I expression and mutations. For example, deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to radiolabeled cytostatin I RNA or alternatively, radiolabeled cytostatin I antisense DNA sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase A digestion or by differences in melting temperatures.

Sequence differences between a reference gene and genes having mutations also may be revealed by direct DNA sequencing. In addition, cloned DNA segments may be employed as probes to detect specific DNA segments. The sensitivity of such methods can be greatly enhanced by appropriate use of PCR or another amplification method. For example, a sequencing primer is used with double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotide or by automatic sequencing procedures with fluorescent-tags.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents. Small sequence deletions and insertions can be visualized by high resolution gel electrophoresis. DNA fragments of different sequences may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (see, e.g., Myers et al., Science, 230: 1242 (1985)).

Sequence changes at specific locations also may be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (e.g., Cotton et al., Proc. Natl. Acad. Sci., USA, 85: 4397-4401 (1985)).

Thus, the detection of a specific DNA sequence may be achieved by methods such as hybridization, RNase protection, chemical cleavage, direct DNA sequencing or the use of restriction enzymes, (e.g., restriction fragment length polymorphisms ("RFLP") and Southern blotting of genomic DNA.

In addition to more conventional gel-electrophoresis and DNA sequencing, mutations also can be detected by in situ analysis.

Chromosome assays

The sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. Moreover, there is a current need for identifying particular sites on the chromosome. Few chromosome marking reagents based on actual sequence data (repeat polymorphisms) are presently available for marking chromosomal location. The mapping of DNAs to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated with disease.

In certain preferred embodiments in this regard, the cDNA herein disclosed is used to clone genomic DNA of a cytostatin I gene. This can be accomplished using a variety of well known techniques and libraries, which generally are available commercially. The genomic DNA the is used for in situ chromosome mapping using well known techniques for this purpose. Typically, in accordance with routine procedures for chromosome mapping, some trial and error may be necessary to identify a genomic probe that gives a good in situ hybridization signal.

In some cases, in addition, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the cDNA. Computer analysis of the 3' untranslated region of the gene is used to rapidly select primers that do not span more than one exon in the genomic DNA, thus

complicating the amplification process. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the primer will yield an amplified fragment.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular DNA to a particular chromosome. Using the present invention with the same oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes or pools of large genomic clones in an analogous manner. Other mapping strategies that can similarly be used to map to its chromosome include in situ hybridization, prescreening with labeled flow-sorted chromosomes and preselection by hybridization to construct chromosome specific-cDNA libraries.

Fluorescence in situ hybridization ("FISH") of a cDNA clone to a metaphase chromosomal spread can be used to provide a precise chromosomal location in one step. This technique can be used with cDNA as short as 50 or 60. For a review of this technique, see Verma et al., HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES, Pergamon Press, New York (1988).

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, MENDELIAN INHERITANCE IN MAN, available on line through Johns Hopkins University, Welch Medical Library. The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritance of physically adjacent genes).

Next, it is necessary to determine the differences in the cDNA or genomic sequence between affected and unaffected individuals. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

With current resolution of physical mapping and genetic mapping techniques, a cDNA precisely localized to a chromosomal region associated with the disease could be one of between 50 and 500 potential causative genes. (This assumes 1 megabase mapping resolution and one gene per 20 kb).

Polypeptide assays

The present invention also relates to a diagnostic assays such as quantitative and diagnostic assays for detecting levels of cytostatin I protein in cells and tissues, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for detecting under-expression of cytostatin I protein compared to normal control tissue samples may be used to detect the presence of aberrant cellular proliferation, for example. Assay techniques that can be used to determine levels of a protein, such as an cytostatin I protein of the present invention, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays. Among these ELISAs frequently are preferred. An ELISA assay initially comprises preparing an antibody specific to cytostatin I, preferably a monoclonal antibody. In addition a reporter antibody generally is prepared which binds to the monoclonal antibody. The reporter antibody is attached a detectable reagent such as radioactive, fluorescent or enzymatic reagent, in this example horseradish peroxidase enzyme.

To carry out an ELISA a sample is removed from a host and incubated on a solid support, e.g. a polystyrene dish, that binds the proteins in the sample. Any free protein binding sites on the dish are then covered by incubating with a non-specific protein such as bovine serum albumin. Next, the monoclonal antibody is incubated in the dish during which time the monoclonal antibodies attach to any cytostatin I proteins attached to the polystyrene dish.

Unbound monoclonal antibody is washed out with buffer. The reporter antibody linked to horseradish peroxidase is placed in the dish resulting in binding of the reporter antibody to any monoclonal antibody bound to cytostatin I. Unattached reporter antibody is then washed out. Reagents for peroxidase activity, including a colorimetric substrate are then added to the dish. Immobilized peroxidase, linked to cytostatin I through the primary and secondary antibodies, produces a colored reaction product. The amount of color developed in a given time period indicates the amount of cytostatin I protein present in the sample. Quantitative results typically are obtained by reference to a standard curve.

A competition assay may be employed wherein antibodies specific to cytostatin I attached to a solid support and labeled cytostatin I and a sample derived from the host are passed over the solid support and the amount of label detected attached to the solid support can be correlated to a quantity of cytostatin I in the sample.

Antibodies

The polypeptides, their fragments or other derivatives, or analogs thereof, or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The present invention also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides corresponding to a sequence of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native

polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., Nature 256: 495-497 (1975), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., Immunology Today 4: 72 (1983) and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole et al., pg. 77-96 in MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc. (1985).

Techniques described for the production of single chain antibodies (U.S. Patent No. 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention. Also, transgenic mice, or other organisms such as other mammals, may be used to express humanized antibodies to immunogenic polypeptide products of this invention.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or purify the polypeptide of the present invention by attachment of the antibody to a solid support for isolation and/or purification by affinity chromatography.

Thus, among others, the medical relevance and practical use of the cytostatin I of the present invention are based upon comparisons with other known proteins and by experimental analysis. The experimental results suggest that cytostatin I, as a therapeutic protein, may have the following medical applications:

1. Anti-tumor: the growth inhibitory activity of cytostatin I may be used as a therapeutic agent to treat various cancers. The expression of cytostatin I in human breast cancers was first investigated in seven breast cancer cell lines: MCF-7, T47D, MDA-MD-231, MDA-MD-435, MDA-MD-436, BT549 and Hs578t. Northern blot analysis failed to detect the cytostatin I transcript in all breast cancer cell lines. The inability to pick up the cytostatin I mRNA in breast cancer cell lines by Northern blot

indicates that the expression of the cytostatin I gene may be down-regulated in breast cancers during the breast malignant progression.

To evaluate the potential biological significance of cytostatin I to human breast cancer progression, we studied cytostatin I gene expression in human breast tumor biopsy samples. The expression of cytostatin I in metastatic breast carcinomas and benign breast tissues were analyzed by Northern blot. Fig. 1 shows a downward progression in the levels of cytostatin I from benign breast to the highly metastatic breast carcinomas. Four of the four RNA samples from benign breast fibroadenomas showed a 1.1 kb transcript. RNAs from sample B5, a breast hyperplasia, showed a very weak 1.1 cytostatin I transcript. In contrast, no signal of the cytostatin I transcript can be detected in all metastatic breast carcinomas except sample C3. The RNA from sample C3 represents an inflammatory breast carcinoma that carried many infiltrating lymphocytes. The strong signal of the cytostatin I transcript may be derived from the infiltrating lymphocytes. The existence of cytostatin I transcripts in benign human breast tissues and its loss of expression in breast carcinomas indicate a role of down-regulation of cytostatin I in breast cancer progression.

In order to localize the cellular source of the cytostatin I expression and to further assess the biological relevance of the down-regulation of cytostatin I expression in breast cancers, *in situ* hybridization was done on fixed sections from 10 in situ ductal carcinomas, 10 infiltrating carcinomas, and 13 benign breast lesions including 7 benign breast fibroadenomas and 6 benign breast hyperplasia (Fig. 2). In these experiments, two aspects of MDGI-1 expression were examined: a) the tissue localization (stromal versus epithelial) of the cytostatin I and b) the correlation of cytostatin I expression and breast malignant phenotype. In all cases a strong cytostatin I transcript was found in the epithelial cells of benign breast fibroadenomas (Fig.2A). The labeling of cytostatin I mRNA

was detectable in the epithelial cells in all seven benign breast fibroadenomas. In contrast, in all cases the highly infiltrating malignant breast samples are not labeled either in the neoplastic cells themselves or their surrounding stromal cells (Fig. 2D). Nine of ten low grade *in situ* carcinomas were also stained negatively (Fig. 2C). In benign breast hyperplasia, five of the six samples showed a negative staining (Fig. 2B) and one sample showed a sparse and a light staining of cytostatin I transcript. The loss of expression in both breast carcinomas and the highly proliferative benign breast hyperplasia (some may eventually become carcinomas) suggest the role of cytostatin I as an anti-proliferative or tumor suppressor gene in breast cancer onset and progression.

2. Anti-angiogenesis: cytostatin I inhibits fibroblast and endothelial cell growth.

3. Anti-metastasis: tumor cells must attract new vessels in order to grow and metastasize efficiently. The inhibition of endothelial cell growth by cytostatin I, therefore, prevents metastases.

4. Stimulation of milk production after childbirth: cytostatin I inhibits mammary epithelial cell growth and modulation mammary gland differentiation, promotes formation of alveolar buds, supports development of differentiated lobuloalveoli, and stimulates milk protein synthesis and fat droplet accumulation.

5. Stimulation of dairy cows milk production or recombinant proteins produced by cows.

6. Modulation of beta-adrenergic sensitivity of cardiac myocytes.

The various potential theapeutic categories and uses of the cytostatin I include but are not limited to all aspects of the following areas of medical practice: 1. Oncology, 2. Cardiovascular, 3. Immunology, 4. Hematology, 5. Metabolism, 6. Gynecology and Obstetrics, and 7. Endocrinology.

Cytostatin I binding molecules and assays

This invention also provides a method for identification of molecules, such as receptor molecules, that bind cytostatin I. Genes encoding proteins that bind cytostatin I, such as receptor proteins, can be identified by numerous methods known to those of skill in the art, for example, ligand panning and FACS sorting. Such methods are described in many laboratory manuals such as, for instance, Coligan et al., Current Protocols in Immunology 1(2): Chapter 5 (1991).

For instance, expression cloning may be employed for this purpose. To this end polyadenylated RNA is prepared from a cell responsive to cytostatin I, a cDNA library is created from this RNA, the library is divided into pools and the pools are transfected individually into cells that are not responsive to cytostatin I. The transfected cells then are exposed to labeled cytostatin I. (Cytostatin I can be labeled by a variety of well-known techniques including standard methods of radio-iodination or inclusion of a recognition site for a site-specific protein kinase.) Following exposure, the cells are fixed and binding of cytostatin I is determined. These procedures conveniently are carried out on glass slides.

Pools are identified of cDNA that produced cytostatin I-binding cells. Sub-pools are prepared from these positives, transfected into host cells and screened as described above. Using an iterative sub-pooling and re-screening process, one or more single clones that encode the putative binding molecule, such as a receptor molecule, can be isolated.

Alternatively a labeled ligand can be photoaffinity linked to a cell extract, such as a membrane or a membrane extract, prepared from cells that express a molecule that it binds, such as a receptor molecule. Cross-linked material is resolved by polyacrylamide gel electrophoresis ("PAGE") and exposed to X-ray film. The labeled complex containing the ligand-receptor can be excised, resolved into peptide fragments, and subjected to protein microsequencing. The amino acid sequence obtained from

microsequencing can be used to design unique or degenerate oligonucleotide probes to screen cDNA libraries to identify genes encoding the putative receptor molecule.

Polypeptides of the invention also can be used to assess cytostatin I binding capacity of cytostatin I binding molecules, such as receptor molecules, in cells or in cell-free preparations.

Agonists and antagonists - assays and molecules

The invention also provides a method of screening compounds to identify those which enhance or block the action of cytostatin I on cells, such as its interaction with cytostatin I-binding molecules such as receptor molecules. An agonist is a compound which increases the natural biological functions of cytostatin I, while antagonists decrease or eliminate such functions.

For example, a cellular compartment, such as a membrane or a preparation thereof, such as a membrane-preparation, may be prepared from a cell that expresses a molecule that binds cytostatin I, such as a molecule of a signaling or regulatory pathway modulated by cytostatin I. The preparation is incubated with labeled cytostatin I in the absence or the presence of a candidate molecule which may be a cytostatin I agonist or antagonist. The ability of the candidate molecule to bind the binding molecule is reflected in decreased binding of the labeled ligand. Molecules which bind gratuitously, i.e., without inducing the effects of cytostatin I on binding the cytostatin I binding molecule, are most likely to be good antagonists. Molecules that bind well and elicit effects that are the same as or closely related to cytostatin I, are good agonists.

Cytostatin I-like effects of potential agonists and antagonists may by measured, for instance, by determining activity of a second messenger system following interaction of the candidate molecule with a cell or appropriate cell preparation, and comparing the effect with that of cytostatin I or molecules that elicit the same effects as

cytostatin I. Second messenger systems that may be useful in this regard include but are not limited to AMP guanylate cyclase, ion channel or phosphoinositide hydrolysis second messenger systems.

Another example of an assay for cytostatin I antagonists is a competitive assay that combines cytostatin I and a potential antagonist with membrane-bound cytostatin I receptor molecules or recombinant cytostatin I receptor molecules under appropriate conditions for a competitive inhibition assay. Cytostatin I can be labeled, such as by radioactivity, such that the number of cytostatin I molecules bound to a receptor molecule can be determined accurately to assess the effectiveness of the potential antagonist.

Potential antagonists include small organic molecules, peptides, polypeptides and antibodies that bind to a polypeptide of the invention and thereby inhibit or extinguish its activity. Potential antagonists also may be small organic molecules, a peptide, a polypeptide such as a closely related protein or antibody that binds the same sites on a binding molecule, such as a receptor molecule, without inducing cytostatin I-induced activities, thereby preventing the action of cytostatin I by excluding cytostatin I from binding.

Other potential antagonists include antisense molecules. Antisense technology can be used to control gene expression through antisense DNA or RNA or through triple-helix formation. Antisense techniques are discussed, for example, in - Okano, J. Neurochem. 56: 560 (1991); OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988). Triple helix formation is discussed in, for instance Lee et al., Nucleic Acids Research 6: 3073 (1979); Cooney et al., Science 241: 456 (1988); and Dervan et al., Science 251: 1360 (1991). The methods are based on binding of a polynucleotide to a complementary DNA or RNA. For example, the 5' coding portion of a polynucleotide that encodes the mature polypeptide of the present invention may be used to design

an antisense RNA oligonucleotide of from about 10 to 40 base pairs in length. A DNA oligonucleotide is designed to be complementary to a region of the gene involved in transcription thereby preventing transcription and the production of cytostatin I. The antisense RNA oligonucleotide hybridizes to the mRNA in vivo and blocks translation of the mRNA molecule into cytostatin I polypeptide. The oligonucleotides described above can also be delivered to cells such that the antisense RNA or DNA may be expressed in vivo to inhibit production of cytostatin I.

The antagonists may be employed in a composition with a pharmaceutically acceptable carrier, e.g., as hereinafter described.

The antagonists may be employed for instance to treat and/or prevent excessive inhibition of cell or tissue growth or inappropriate differentiation stimulatory activity. For example, the antagonists promote involution of breast (return of an enlarged breast to normal size after parturition, childbirth): Antisense phosphorothioate oligonucleotides or antibodies to cytostatin I could selectively inhibit endogenous cytostatin I expression in mammary epithelial cells and suppresses appearance of alveolar end buds and lowers the beta-casein level.

Compositions

The invention also relates to compositions comprising the polynucleotide or the polypeptides discussed above or the agonists or antagonists. Thus, the polypeptides of the present invention may be employed in combination with a non-sterile or sterile carrier or carriers for use with cells, tissues or organisms, such as a pharmaceutical carrier suitable for administration to a subject. Such compositions comprise, for instance, a media additive or a therapeutically effective amount of a polypeptide of the invention and a pharmaceutically acceptable carrier or excipient. Such carriers may include, but are not limited to, saline, buffered saline, dextrose, water, glycerol,

ethanol and combinations thereof. The formulation should suit the mode of administration.

Kits

The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, reflecting approval by the agency of the manufacture, use or sale of the product for human administration.

Administration

Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

The pharmaceutical compositions may be administered in any effective, convenient manner including, for instance, administration by topical, oral, anal, vaginal, intravenous, intraperitoneal, intramuscular, subcutaneous, intranasal or intradermal routes among others.

The pharmaceutical compositions generally are administered in an amount effective for treatment or prophylaxis of a specific indication or indications. In general, the compositions are administered in an amount of at least about 10 μg/kg body weight. In most cases they will be administered in an amount not in excess of about 8 mg/kg body weight per day. Preferably, in most cases, dose is from about 10 μg/kg to about 1 mg/kg body weight, daily. It will be appreciated that optimum dosage will be determined by standard methods for each treatment modality and indication, taking into account the indication, its severity, route of administration, complicating conditions and the like.

Gene therapy

The cytostatin I polynucleotides, polypeptides, agonists and antagonists that are polypeptides may be employed in accordance with the present invention by expression of such polypeptides in vivo, in treatment modalities often referred to as "gene therapy."

Thus, for example, cells from a patient may be engineered with a polynucleotide, such as a DNA or RNA, encoding a polypeptide ex vivo, and the engineered cells then can be provided to a patient to be treated with the polypeptide. For example, cells may be engineered ex vivo by the use of a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention. Such methods are well-known in the art and their use in the present invention will be apparent from the teachings herein.

Similarly, cells may be engineered in vivo for expression of a polypeptide in vivo by procedures known in the art. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct then may be isolated and introduced into a packaging cell is transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a patient for engineering cells in vivo and expression of the polypeptide in vivo. These and other methods for administering a polypeptide of the present invention by such method should be apparent to those skilled in the art from the teachings of the present invention.

Retroviruses from which the retroviral plasmid vectors herein above mentioned may be derived include, but are not limited to, Moloney Murine Leukemia Virus, spleen necrosis virus, retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, gibbon ape leukemia virus, human immunodeficiency virus, adenovirus, Myeloproliferative Sarcoma Virus, and mammary tumor virus.

In one embodiment, the retroviral plasmid vector is derived from Moloney Murine Leukemia Virus.

Such vectors well include one or more promoters for expressing the polypeptide. Suitable promoters which may be employed include, but are not limited to, the retroviral LTR; the SV40 promoter; and the human cytomegalovirus (CMV) promoter described in Miller et al., Biotechniques 7: 980-990 (1989), or any other promoter (e.g., cellular promoters such as eukaryotic cellular promoters including, but not limited to, the histone, RNA polymerase III, and ß-actin promoters). Other viral promoters which may be employed include, but are not limited to, adenovirus promoters, thymidine kinase (TK) promoters, and B19 parvovirus promoters. The selection of a suitable promoter will be apparent to those skilled in the art from the teachings contained herein.

The nucleic acid sequence encoding the polypeptide of the present invention will be placed under the control of a suitable promoter. Suitable promoters which may be employed include, but are not limited to, adenoviral promoters, such as the adenoviral major late promoter; or heterologous promoters, such as the cytomegalovirus (CMV) promoter; the respiratory syncytial virus (RSV) promoter; inducible promoters, such as the MMT promoter, the metallothionein promoter; heat shock promoters; the albumin promoter; the ApoAI promoter; human globin promoters; viral thymidine kinase promoters, such as the Herpes Simplex thymidine kinase promoter; retroviral LTRs (including the modified retroviral LTRs herein above described); the ß-actin promoter; and human growth hormone promoters. The promoter also may be the native promoter which controls the gene encoding the polypeptide.

The retroviral plasmid vector is employed to transduce packaging cell lines to form producer cell lines. Examples of packaging cells which may be transfected include, but are not limited to, the PE501, PA317, Y-2, Y-AM, PA12, T19-14X, VT-19-17-H2, YCRE, YCRIP, GP+E-86, GP+envAm12, and DAN cell lines as described in Miller, A., Human Gene Therapy

1: 5-14 (1990). The vector may be transduced into the packaging cells through any means known in the art. Such means include, but are not limited to, electroporation, the use of liposomes, and CaPO4 precipitation. In one alternative, the retroviral plasmid vector may be encapsulated into a liposome, or coupled to a lipid, and then administered to a host.

The producer cell line will generate infectious retroviral vector particles, which include the nucleic acid sequence(s) encoding the polypeptides. Such retroviral vector particles then may be employed to transduce eukaryotic cells, either in vitro or in vivo. The transduced eukaryotic cells will express the nucleic acid sequence(s) encoding the polypeptide. Eukaryotic cells which may be transduced include, but are not limited to, embryonic stem cells, embryonic carcinoma cells, as well as hematopoietic stem cells, hepatocytes, fibroblasts, myoblasts, keratinocytes, endothelial cells, and bronchial epithelial cells.

EXAMPLES

The present invention is further described by the following examples. The examples are provided solely to illustrate the invention by reference to specific embodiments. These exemplification's, while illustrating certain specific aspects of the invention, do not portray the limitations or circumscribe the scope of the disclosed invention.

Certain terms used herein are explained in the foregoing glossary.

All examples were carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. Routine molecular biology techniques of the following examples can be carried out as described in standard laboratory manuals, such as Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed.; Cold

Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), herein referred to as "Sambrook."

All parts or amounts set out in the following examples are by weight, unless otherwise specified.

Unless otherwise stated size separation of fragments in the examples below was carried out using standard techniques of agarose and polyacrylamide gel electrophoresis ("PAGE") in Sambrook and numerous other references such as, for instance, by Goeddel et al., Nucleic Acids Res. 8: 4057 (1980).

Unless described otherwise, ligations were accomplished using standard buffers, incubation temperatures and times, approximately equimolar amounts of the DNA fragments to be ligated and approximately 10 units of T4 DNA ligase ("ligase") per 0.5 μg of DNA.

Example 1

Expression and purification of human cytostatin I using bacteria

The DNA sequence encoding human cytostatin I in the deposited polynucleotide was amplified using PCR oligonucleotide primers specific to the amino and carboxyl terminal sequence of the human cytostatin I protein and to vector sequences 3′ to the gene. Additional nucleotides containing restriction sites to facilitate cloning were added to the 5′ and 3′ sequences respectively.

The 5′ oligonucleotide primer had the sequence 5′ CGCGGATCCATGCCTCCCAACCTCACTG 3′ containing the underlined BamHI restriction site, which encodes a start AUG, followed by 19 nucleotides of the human cytostatin I coding sequence beginning with the starting codon of the gene.

The 3′ primer had the sequence 5′ GCGTCTAGACT ATCTGACCTTCCTGAAGAC3′ containing the underlined XbaI site restriction site followed by 20 nucleotides of cytostatin I including the stop codon.

The restrictions sites were convenient to restriction enzyme sites in the bacterial expression vectors pQE-9

which were used for bacterial expression in these examples. (Qiagen, Inc. Chatsworth, CA). pQE-9 encodes ampicillin antibiotic resistance ("Ampr") and contains a bacterial origin of replication ("ori"), an IPTG inducible promoter, a ribosome binding site ("RBS"), a 6-His tag and restriction enzyme sites.

The amplified human cytostatin I DNA and the vector pQE-9 both were digested with BamHI and XbaI and the digested DNAs then were ligated together. Insertion of the cytostatin I DNA into the pQE-9 restricted vector placed the cytostatin I coding region downstream of and operably linked to the vector's IPTG-inducible promoter and in-frame with an initiating AUG appropriately positioned for translation of cytostatin I.

The ligation mixture was transformed into competent E. coli M15/rep4 cells using standard procedures. Such procedures are described in Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd Ed.; Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989). E. coli strain M15/rep4, containing multiple copies of the plasmid pREP4, which expresses lac repressor and confers kanamycin resistance ("Kanr"), was used in carrying out the illustrative example described here. This strain, which is only one of many that are suitable for expressing cytostatin I, is available commercially from Qiagen.

Transformants were identified by their ability to grow on LB plates in the presence of ampicillin. Plasmid DNA was isolated from resistant colonies and the identity of the cloned DNA was confirmed by restriction analysis.

Clones containing the desired constructs were grown overnight ("O/N") in liquid culture in LB media supplemented with both ampicillin (100 ug/ml) and kanamycin (25 ug/ml).

The O/N culture was used to inoculate a large culture, at a dilution of approximately 1:100 to 1:250. The cells were grown to an optical density at 600nm ("OD600") of between 0.4 and 0.6. Isopropyl-B-D-thiogalactopyranoside ("IPTG") was then added to a final concentration of 1 mM to

induce transcription from lac repressor sensitive promoters, by inactivating the lacI repressor. Cells subsequently were incubated further for 3 to 4 hours. Cells then were harvested by centrifugation and disrupted, by standard methods. The cell pellet was solubilized in the chaotropic agent 6 Molar Guanidine HCl. After clarification, solubilized cytostatin I was purified from this solution by chromatography on a Nickel-Chelate column under conditions that allow for tight binding by proteins containing the 6-His tag (Hochuli, E. et al., J. Chromatography 411:177-184 (1984)). Cytostatin I (90 % pure was eluted from the column in 6 molar guanidine HCl pH 5.0 and for the purpose of renaturation adjusted to 3 molar guanidine HCl, 100mM sodium phosphate, 10 mmolar glutathione (reduced) and 2 mmolar glutathione (oxidized). After incubation in this solution for 12 hours the protein was dialyzed to 10 mmolar sodium phosphate.

The entire coding sequence including the putative signal sequence or transmembrane domain was fused in frame with a 6-His tag present in the expression vector pQE9. E. coli harboring the expression plasmid were induced with 1 mM IPTG during the logarithmic growth phase. Following a 3-hour induction, the cell pellet was lysed with 6M Guanidine hydrochloride and cytostatin I was purified using a Nickel-chelate affinity chromatography column. The highly purified protein was denatured by dialysis in PBS buffer. The gel is shown in Figure 4: M, molecular weight markers; Lane 1 and 2, induced cell lysate; Lane 3 and 4, uninduced cell lysate; Lane 5, pass through fraction from Nickel-chelate column purification; Lane 6, 7 and 8, Fraction eluted with 6M Guanidine hydrochloride (pH 5); 9 Fraction eluted with 6M Guanidine hydrochloride (pH 2).

Example 2

Cloning and expression of human cytostatin I in a baculovirus expression system

The cDNA sequence encoding the full length human cytostatin I protein, in the deposited clone is amplified

using PCR oligonucleotide primers corresponding to the 5' and 3' sequences of the gene:

The 5' primer has the sequence 5'CGC GGA TCC CCC TCC CAA CCT CAC TGG CTA C 3' containing the underlined BamHI restriction enzyme site followed by 22 nucleotides of the sequence of cytostatin I of Figure 1. Inserted into an expression vector, as described below, the 5' end of the amplified fragment encoding human cytostatin I provides an efficient signal peptide. An efficient signal for initiation of translation in eukaryotic cells, as described by Kozak, M., J. Mol. Biol. 196: 947-950 (1987) is appropriately located in the vector portion of the construct.

The 3' primer has the sequence 5' CGC GGA TCC CTA TCT GAC CTT CCT GAA GA 3' containing the underlined BamHI restriction followed by 20 nucleotides of the C-terminal cytostatin I coding sequence set out in Figure 1, including the stop codon.

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("Geneclean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with BamH1 and Asp718 and again is purified on a 1% agarose gel. This fragment is designated herein F2.

The vector pA2-Gp is used to express the cytostatin I protein in the baculovirus expression system, using standard methods, such as those described in Summers et al, A MANUAL OF METHODS FOR BACULOVIRUS VECTORS AND INSECT CELL CULTURE PROCEDURES, Texas Agricultural Experimental Station Bulletin No. 1555 (1987). This expression vector contains the strong polyhedrin promoter of the Autographa californica nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites. The signal peptide of AcMNPV gp67, including the N-terminal methionine, is located just upstream of a BamH1 site. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For an easy selection of recombinant virus the beta-galactosidase gene from E.coli is inserted in the same orientation as the polyhedrin promoter and is

followed by the polyadenylation signal of the polyhedrin gene. The polyhedrin sequences are flanked at both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate viable virus that express the cloned polynucleotide.

Many other baculovirus vectors could be used in place of pA2-GP, such as pAc373, pVL941 and pAcIM1 provided, as those of skill readily will appreciate, that construction provides appropriately located signals for transcription, translation, trafficking and the like, such as an in-frame AUG and a signal peptide, as required. Such vectors are described in Luckow et al., Virology 170: 31-39, among others.

The plasmid is digested with the restriction enzyme BamHI and then is dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("Geneclean" BIO 101 Inc., La Jolla, Ca.). This vector DNA is designated herein "V2".

Fragment F2 and the dephosphorylated plasmid V2 are ligated together with T4 DNA ligase. E.coli HB101 cells are transformed with ligation mix and spread on culture plates. Bacteria are identified that contain the plasmid with the human cytostatin I gene by digesting DNA from individual colonies using BamHI and then analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing. This plasmid is designated herein pBaccytostatin I.

5 μg of the plasmid pBaccytostatin I is co-transfected with 1.0 μg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA.), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84: 7413-7417 (1987). 1μg of BaculoGold™ virus DNA and 5 μg of the plasmid pBaccytostatin I are mixed in a sterile well of a microtiter plate containing 50 μl of serum free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards 10 μl Lipofectin plus 90 μl Grace's medium are

added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is rocked back and forth to mix the newly added solution. The plate is then incubated for 5 hours at 27°C. After 5 hours the transfection solution is removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. The plate is put back into an incubator and cultivation is continued at 27°C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, cited above. An agarose gel with "Blue Gal" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10).

Four days after serial dilution, the virus is added to the cells. After appropriate incubation, blue stained plaques are picked with the tip of an Eppendorf pipette. The agar containing the recombinant viruses is then resuspended in an Eppendorf tube containing 200 μl of Grace's medium. The agar is removed by a brief centrifugation and the supernatant containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4°C. A clone containing properly inserted cytostatin I is identified by DNA analysis including restriction mapping and sequencing. This is designated herein as V-cytostatin I.

Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus V-cytostatin I at a

multiplicity of infection ("MOI") of about 2 (about 1 to about 3). Six hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Gaithersburg). 42 hours later, 5 μCi of 35S-methionine and 5 μCi 35S cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then they are harvested by centrifugation, lysed and the labeled proteins are visualized by SDS-PAGE and autoradiography.

Example 3

Expression of Recombinant Cytostatin I in COS cells

The expression of plasmid containing the cytostatin I gene is derived from a vector pcDNAI/Amp (Invitrogen) containing: 1) SV40 origin of replication, 2) ampicillin resistance gene, 3) E.coli replication origin, 4) CMV promoter followed by a polylinker region, an SV40 intron and polyadenylation site. A DNA fragment encoding the entire cytostatin I precursor and a HA tag fused in frame to its 3' end is cloned into the polylinker region of the vector, therefore, the recombinant protein expression is directed under the CMV promoter. The HA tag corresponds to an epitope derived from the influenza hemagglutinin protein as previously described (I. Wilson, H. Niman, R. Heighten, A Cherenson, M. Connolly, and R. Lerner, 1984, Cell 37:767, (1984)). The infusion of HA tag to the target protein allows easy detection of the recombinant protein with an antibody that recognizes the HA epitope.

The plasmid construction strategy is described as follows:

The DNA sequence encoding cytostatin I, ATCC is constructed by PCR on the original cytostatin I cloned using two primers: the 5' primer from the 5' end of the cytostatin I gene and a 3' sequence from the 3' end of the cytostatin I gene. Therefore, the PCR product contains the a cytostatin I coding sequence followed by HA tag fused in frame, a translation termination stop codon next to the HA tag, and a final restriction endonuclease site. The PCR

amplified DNA fragment and the vector, pcDNAI/Amp, are digested with the appropriate restriction enzymes and ligated. The ligation mixture is transformed into E. coli strain SURE (available from Stratagene Cloning Systems, 11099 North Torrey Pines Road, La Jolla, CA 92037) the transformed culture is plated on ampicillin media plates and resistant colonies are selected. Plasmid DNA is isolated from transformants and examined by restriction analysis for the presence of the correct fragment. For expression of the recombinant cytostatin I, COS cells are transfected with the expression vector by DEAE-DEXTRAN method (J. Sambrook, E. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, Cold Spring Laboratory Press, (1989)). The expression of the cytostatin I HA protein is detected by radiolabelling and immunoprecipitation method (E. Harlow, D. Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press, (1988)). Cells are labelled for 9 hours with 35S-cysteine two days post transfection. Culture media is then collected and cells are lysed with detergent (RIPA buffer (150 MM NaCl, 1% NP-40, 0.1% SDS, 1% NP-40, 0.5% DOC, 50mM Tris, pH 7.5) (Wilson, I. et al., Id. 37:767 (1984)). Both cell lysate and culture media are precipitated with an HA specific monoclonal antibody. Proteins precipitated are analyzed on 15% SDS-PAGE gels.

Example 4

Expression pattern of cytostatin I in human tissue

Northern blot analysis is carried out to examine the levels of expression of cytostatin I in human tissues. Total cellular RNA samples are isolated with RNAzol® B system (Biotecx Laboratories, Inc. 6023 South Loop East, Houston, TX 77 03 3) . About 10[ig of total RNA isolated from each human tissue specified is separated on 1% agarose gel and blotted onto a nylon filter (Sambrook, Fritsch, and Maniatis, Molecular Cloning, Cold Spring Harbor Press, (1989)). The labeling reaction is done according to the Stratagene PrimeIt kit with 50ng DNA fragment. The labeled DNA is purified with a Select-G-50 column (5 Prime - 3

Prime, Inc. 5603 Arapahoe Road, Boulder, CO 80303) . The filter is then hybridized with radioactive labeled full length cytostatin I gene at 1,000,000 cpm/ml in 0. 5 M NaPO4, pH 7. 4 and 7% SDS overnight at 65°C. Af ter washing twice at room temperature and twice at 60°C with 0.5 x SSC, 0.1% SDS, the filter is then exposed at -70°C overnight with an intensifying screen. Figure 3A issustrates the tissue distribution of cytostatin I in various human tissues. The results are issustrated in figures 3A, 3B and 3C.

Example 5

Gene therapeutic expression of human cytostatin I

Fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature overnight. After 24 hours at room temperature, the flask is inverted - the chunks of tissue remain fixed to the bottom of the flask - and fresh media is added (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin). The tissue is then incubated at 37°C for approximately one week. At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerges. The monolayer is trypsinized and scaled into larger flasks.

A vector for gene therapy is digested with restriction enzymes for cloning a fragment to be expressed. The digested vector is treated with calf intestinal phosphatase to prevent self-ligation. The dephosphorylated, linear vector is fractionated on an agarose gel and purified.

Cytostatin I cDNA capable of expressing active cytostatin I, is isolated. The ends of the fragment are modified, if necessary, for cloning into the vector. For instance, 5" overhanging may be treated with DNA

polymerase to create blunt ends. 3' overhanging ends may be removed using S1 nuclease. Linkers may be ligated to blunt ends with T4 DNA ligase.

Equal quantities of the Moloney murine leukemia virus linear backbone and the cytostatin I fragment are mixed together and joined using T4 DNA ligase. The ligation mixture is used to transform E. Coli and the bacteria are then plated onto agar-containing kanamycin. Kanamycin phenotype and restriction analysis confirm that the vector has the properly inserted gene.

Packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The vector containing the cytostatin I gene is introduced into the packaging cells by standard techniques. Infectious viral particles containing the cytostatin I gene are collected from the packaging cells, which now are called producer cells.

Fresh media is added to the producer cells, and after an appropriate incubation period media is harvested from the plates of confluent producer cells. The media, containing the infectious viral particles, is filtered through a Millipore filter to remove detached producer cells. The filtered media then is used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the filtered media. Polybrene (Aldrich) may be included in the media to facilitate transduction. After appropriate incubation, the media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his, to select out transduced cells for expansion.

Engineered fibroblasts then may be injected into rats, either alone or after having been grown to confluence on microcarrier beads, such as cytodex 3 beads. The injected

fibroblasts produce cytostatin I product, and the biological actions of the protein are conveyed to the host.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples.

Example 6

Biological Activity of Cytostatin I

The activity of cytostatin I is illustrated in Figure 5. Two-fold serial dilution of purified cytostatin I (MDGI homolog, HGO7400-1E or HGO7400-2E) starting from 100 ng/ml were made in RPMI 1640 medium with 0.5% FBS. The adherent target cells were prepared from confluent cultures by trypsinization in PBS, and non-adherent target cells were harvested from stationary cultures and washed once with medium. Target cells were suspended at 1 x 10^5 cells/ml in medium containing 0.5% FBS, then 0.1 ml aliquots were dispensed into 96-well flat-bottomed microtiter plates containing 0.1 ml serially diluted test samples. Incubation was continued for 70 hr. The activity was quantified using MTS [3(4,5-dimethyl-thiazoyl-2-yl) 5 (3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium)] Assay. MTS assay is performed by the addition of 20 til of MTS and phenazine methosulfate (PMS) solution to 96 well plates (Stock solution was prepared as described by Promega Technical Bulletin No. 169). During a 3 hr incubation, living cells convert the MTS into a the aqueous soluble formazan product. Wells with medium only (no cells) were processed in exactly the same manner as the rest of the wells and were used for blank controls. Wells with medium and cells were used as baseline controls. The absorbence at 490 nm was recorded using an ELISA reader and is proportional to the number of viable cells in the wells. Cell growth promotion (positive percentage) or inhibition (negative percentage), as a percentage compared to baseline control wells (variation between three baseline control well is less than 5%), calculated for each sample concentration, by the formula: $OD_{experimental}/OD_{baseline\ control}$ X 100

-100. All determinations were made in triplicate. Mean and SD were calculated by Microsoft Excel.

Example 7

In situ Hybridization Conditions

Deparaffinized and acid-treated sections (5-um thick) were treated with proteinase K (0.2 mg/ml) for 30 min, prehybridized at 50°C for 4 hours, and hybridized overnight with digoxigenin labeled anti-sense transcripts from a TIMP-4 or TIMP-2 cDNA insert. The TIMP-4 antisense transcript is a 390 bp riboprobe as described for Northern blot. The TIMP-2 probe is a 1.1 kb antisense probe that was generated from NotI-digested TIMP-2 template. Hybridization (50°C for 18 hours) followed by RNase treatment (40 ug/ml, 3 0 min at 37°C) and three stringent washings (60°C for 40 min). Sections were incubated with mouse anti-digoxigenin antibodies (Boehringer) followed by the incubation with biotin-conjugated secondary rabbit anti-mouse antibodies (DAKO). The colorimetric detection were perforMed using a standard indirect streptavidin-biotin immunoreaction method by DAKO's Universal LSAB Kit according to manufacturer's instructions.

Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Ni, Jian
Gentz, Reiner
Yu, Guo-Liang
Rosen, Craig A
Su, Jeffrey

(ii) TITLE OF INVENTION: Human G-Protein Coupled Receptor

(iii) NUMBER OF SEQUENCES: 11

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI, STEWART & OLSTEIN
(B) STREET: 6 BECKER FARM ROAD
(C) CITY: ROSELAND
(D) STATE: NEW JERSEY
(E) COUNTRY: USA
(F) ZIP: 07068-1739

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: Herewith
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Ferraro, Gregory D
(B) REGISTRATION NUMBER: 36,134
(C) REFERENCE/DOCKET NUMBER: 325800-550

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 201-994-1700
(B) TELEFAX: 201-994-1744

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 861 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 94..414

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CACGAGCTGG AATCTCTCAG CCTCACCTGC CAGACAACAC CCCCTCCTTC CTCACCCTGT      60

TTCCTGCATT CTCCTGAAAC CTTCATCCAC ACA ATG CCT CCC AAC CTC ACT GGC      114
                                     Met Pro Pro Asn Leu Thr Gly
```

```
                                                1               5

TAC TAC CGC TTT GTT TCG CAG AAG AAC ATG GAG GAC TAC CTG CAA GCC        162
Tyr Tyr Arg Phe Val Ser Gln Lys Asn Met Glu Asp Tyr Leu Gln Ala
         10                  15                  20

CTA AAC ATC AGC TTG GCT GTG CGG AAG ATC GCG CTG CTG CTG AAG CCG        210
Leu Asn Ile Ser Leu Ala Val Arg Lys Ile Ala Leu Leu Leu Lys Pro
     25                  30                  35

GAC AAG GAG ATC GAA CAC CAG GGC AAC CAC ATG ACG GTG AGG ACG CTC        258
Asp Lys Glu Ile Glu His Gln Gly Asn His Met Thr Val Arg Thr Leu
 40                  45                  50                  55

AGC ACC TTC CGA AAC TAC ACT TTG CAG TTT GAT GTG GGA GTG CAG AAA        306
Ser Thr Phe Arg Asn Tyr Thr Leu Gln Phe Asp Val Gly Val Gln Lys
             60                  65                  70

GGG GAG GTC CCC AAC CGG GGC TGG AGA CAC TGG CTG GAG GGA GAG TTG        354
Gly Glu Val Pro Asn Arg Gly Trp Arg His Trp Leu Glu Gly Glu Leu
         75                  80                  85

CTG TAT CTG GAA CTG ACT GCA AGG GAT GCA GTG TGC GAG CAG GTC TTC        402
Leu Tyr Leu Glu Leu Thr Ala Arg Asp Ala Val Cys Glu Gln Val Phe
         90                  95                 100

AGG AAG GTC AGA TAGCCGGAGA GGAGCCAAGA TCCCTCCAGA CAGCACCAGC            454
Arg Lys Val Arg
    105

TCACAGACGC TCTTGTTGTG CCCCCTTCAA GCCCAGATTG TGCCAGGTCA GCTGTCCCTT      514

CCTCTGGCCA CCTTTCCTCC CTCTGGGTCC CTCCTCACCC CTCCCCGTGT TAATCTGTAA      574

CTTGGAGCCC CCAGGACAAA GTCCTTTCTC ACACTCCACT GCCCAATAGT GACCTCACTT      634

CCAGGTCAAG GTCTGGCGTC CCAAATGAAA GAAGCAGGCA AAGGGAAGGA GCCCCTGAGG      694

ACAACCAATC TCCGCTCTCT CCTGTCCATT TGACCTCTTC TTTTCCTTCT AAGAAAGAAC      754

TAAGCTTTGG GCATTTGGCG ATTAGTGAAA ATTCTATCCT GATGGACTTC TGGAAAACTG      814

TGACTGGGGT TCAACAGTTT AAACAGGGGC TACTGGGGGA AAAAAAA                    861
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 107 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Pro Pro Asn Leu Thr Gly Tyr Tyr Arg Phe Val Ser Gln Lys Asn
  1               5                  10                  15

Met Glu Asp Tyr Leu Gln Ala Leu Asn Ile Ser Leu Ala Val Arg Lys
             20                  25                  30

Ile Ala Leu Leu Leu Lys Pro Asp Lys Glu Ile Glu His Gln Gly Asn
         35                  40                  45
```

His Met Thr Val Arg Thr Leu Ser Thr Phe Arg Asn Tyr Thr Leu Gln
50 55 60

Phe Asp Val Gly Val Gln Lys Gly Glu Val Pro Asn Arg Gly Trp Arg
65 70 75 80

His Trp Leu Glu Gly Glu Leu Leu Tyr Leu Glu Leu Thr Ala Arg Asp
85 90 95

Ala Val Cys Glu Gln Val Phe Arg Lys Val Arg
100 105

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CGCGGATCCA TGCCTCCCAA CCTCACTG 28

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GCGTCTAGAC TATCTGACCT TCCTGAAGAC 30

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGCGGATCCC CCTCCCAACC TCACTGGCTA C 31

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CGCGGATCCC TATCTGACCT TCCTGAAGA 29

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 133 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
Met Ala Asp Ala Phe Val Gly Thr Trp Lys Leu Val Asp Ser Lys Asn
1               5                   10                  15

Phe Asp Asp Tyr Met Lys Ser Leu Gly Val Gly Phe Ala Thr Arg Gln
            20                  25                  30

Val Ala Ser Met Thr Lys Pro Thr Thr Ile Ile Glu Lys Asn Gly Asp
        35                  40                  45

Thr Ile Thr Ile Lys Thr Gln Ser Thr Phe Lys Asn Thr Glu Ile Asn
    50                  55                  60

Phe Gln Leu Gly Ile Glu Phe Asp Glu Val Thr Ala Asp Asp Arg Lys
65                  70                  75                  80

Val Lys Ser Leu Val Thr Leu Asp Gly Gly Lys Leu Ile His Val Gln
                85                  90                  95

Lys Trp Asn Gly Gln Glu Thr Thr Leu Thr Arg Glu Leu Val Asp Gly
            100                 105                 110

Lys Leu Ile Leu Thr Leu Thr His Gly Ser Val Val Ser Thr Arg Thr
        115                 120                 125

Tyr Glu Lys Glu Ala
    130
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 135 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Pro Val Asp Phe Thr Gly Tyr Trp Lys Met Leu Val Asn Glu Asn
1 5 10 15

Phe Glu Glu Tyr Leu Arg Ala Leu Asp Val Asn Val Ala Leu Arg Lys
20 25 30

Ile Ala Asn Leu Leu Lys Pro Asp Lys Glu Ile Val Gln Asp Gly Asp
35 40 45

His Met Ile Ile Arg Thr Leu Ser Thr Phe Arg Asn Tyr Ile Met Asp
50 55 60

Phe Gln Val Gly Lys Glu Phe Glu Glu Asp Leu Thr Gly Ile Asp Asp
65 70 75 80

Arg Lys Cys Met Thr Thr Val Ser Trp Asp Gly Asp Lys Leu Gln Cys
85 90 95

Val Gln Lys Gly Glu Lys Glu Gly Arg Gly Trp Thr Gln Trp Ile Glu
100 105 110

Gly Asp Glu Leu His Leu Glu Met Arg Val Glu Gly Val Val Cys Lys
115 120 125

Gln Val Phe Lys Lys Val Gln
130 135

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 134 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met Thr Arg Asp Gln Asn Gly Thr Trp Glu Met Glu Ser Asn Glu Asn
1 5 10 15

Phe Glu Gly Tyr Met Lys Ala Leu Asp Ile Asp Phe Ala Thr Pro Lys
20 25 30

Ile Ala Val Arg Leu Thr Gln Thr Lys Val Ile Asp Gln Asp Gly Asp
35 40 45

Asn Phe Lys Thr Lys Thr Thr Ser Thr Phe Arg Asn Tyr Asp Val Asp
50 55 60

Phe Thr Val Gly Val Glu Phe Asp Glu Tyr Thr Lys Ser Leu Asp Asn
65 70 75 80

Arg His Val Lys Ala Leu Val Thr Trp Glu Gly Asp Val Leu Val Cys

85 90 95

Val Gln Lys Gly Glu Lys Glu Asn Arg Gly Trp Lys Gln Trp Ile Glu
100 105 110

Gly Asp Lys Leu Tyr Leu Glu Leu Thr Cys Gly Asp Gln Val Cys Arg
115 120 125

Gln Val Phe Lys Lys Lys
130

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 133 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Val Asp Ala Phe Leu Gly Thr Trp Lys Leu Val Asp Ser Lys Asn
1 5 10 15

Phe Asp Asp Tyr Met Lys Ser Leu Gly Val Gly Phe Ala Thr Arg Gln
20 25 30

Val Ala Ser Met Thr Lys Pro Thr Thr Ile Ile Glu Lys Asn Gly Asp
35 40 45

Ile Leu Thr Leu Lys Thr His Ser Thr Phe Lys Asn Thr Glu Ile Ser
50 55 60

Phe Lys Leu Gly Val Glu Phe Asp Glu Thr Thr Ala Asp Asp Arg Lys
65 70 75 80

Val Lys Ser Ile Val Thr Leu Asp Gly Gly Lys Leu Val His Leu Gln
85 90 95

Lys Trp Asp Gly Gln Glu Thr Thr Leu Val Arg Glu Leu Ile Asp Gly
100 105 110

Lys Leu Ile Leu Thr Leu Thr His Gly Thr Ala Val Cys Thr Arg Thr
115 120 125

Tyr Glu Lys Glu Ala
130

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 132 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
Met Ser Asn Lys Phe Leu Gly Thr Trp Lys Leu Val Ser Ser Glu Asn
1               5                   10                  15

Phe Asp Asp Tyr Met Lys Ala Leu Gly Val Gly Leu Ala Thr Arg Lys
            20                  25                  30

Leu Gly Asn Leu Ala Lys Pro Thr Val Ile Ile Ser Lys Lys Gly Asp
        35                  40                  45

Ile Ile Thr Ile Arg Thr Glu Ser Thr Phe Lys Asn Thr Glu Ile Ser
    50                  55                  60

Phe Lys Leu Gly Gln Glu Phe Glu Glu Thr Thr Ala Asp Asn Arg Lys
65                  70                  75                  80

Thr Lys Ser Ile Val Thr Leu Gln Arg Gly Ser Leu Asn Gln Val Gln
                85                  90                  95

Arg Trp Asp Gly Lys Glu Thr Thr Ile Lys Arg Lys Leu Val Asn Gly
            100                 105                 110

Lys Met Val Ala Glu Cys Lys Met Lys Gly Val Val Cys Thr Arg Ile
        115                 120                 125

Tyr Glu Lys Val
    130
```

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO:2;

(b) a polynucleotide having at least a 70% identity to a polynucleotide encoding a mature cytostatin I polypeptide;

(c) a polynucleotide which is complementary to the polynucleotide of (a) or (b); and

(d) a polynucleotide comprising at least 15 bases of the polnucleotide of (a), (b) or (c).

2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.

3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.

4. The polynucleotide of Claim 1 wherein the polynucleotide is genomic DNA.

5. The polynucleotide of Claim 2 which encodes the polypeptide comprising amino acid 1 to 107 of SEQ ID NO:2.

6. The polynucleotide of Claim 2 which encodes the mature cytostatin I polypeptide.

7. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) a polynucleotide which encodes a mature polypeptide having the amino acid sequence expressed by the human cDNA contained in ATCC Deposit No. 97103;

(b) a polynucleotide which is complementary to the polynucleotide of (a); and

(c) a polynucleotide comprising at least 15 bases of the polynucleotide of (a) or (b).

8. The polynucleotide of claim 1 comprising the sequence as set forth in SEQ ID NO:1 from nucleotide 1 to nucleotide 861.

9. The polynucleotide of claim 1 comprising the sequence as set forth in SEQ ID NO:1 from nucleotide 94 to nucleotide 414.

10. The polynucleotide of claim 1 comprising the sequence encoding a mature cytostatin I polypeptide.

11. A vector comprising the DNA of Claim 2.

12. A host cell comprising the vector of Claim 11.

13. A process for producing a polypeptide comprising: expressing from the host cell of Claim 12 the polypeptide encoded by said DNA.

14. A process for producing a cell which expresses a polypeptide comprising genetically engineering cells with the vector of Claim 11.

15. A polypeptide comprising a member selected from the group consisting of:

(a) a polypeptide having an amino acid sequence set forth in SEQ ID NO:2; and

(b) a mature cytostatin I polypeptide; and

(c) a polypeptide which is at least 70% identical to the polypeptide of (a) or (b).

16. A compound which inhibits activation of the polypeptide of claim 15.

17. A compound which activates the polypeptide of claim 15.

18. A method for the treatment of a patient having need of cytostatin I comprising: administering to the patient a therapeutically effective amount of the polypeptide of claim 15.

19. The method of Claim 18 wherein said therapeutically effective amount of the polypeptide is administered by providing to the patient DNA encoding said polypeptide and expressing said polypeptide *in vivo*.

20. A method for the treatment of a patient having need to inhibit a cytostatin I polypeptide comprising: administering to the patient a therapeutically effective amount of the compound of Claim 16.

21. A process for diagnosing a disease or a susceptibility to a disease related to an under-expression of the polypeptide of claim 15 comprising:

determining a mutation in a nucleic acid sequence encoding said polypeptide.

22. A diagnostic process comprising:

analyzing for the presence of the polypeptide of claim 15 in a sample derived from a host.

23. A method for identifying compounds which bind to and inhibit activation of the polypeptide of claim 15 comprising: contacting a cell expressing on the surface thereof a receptor for the polypeptide, said receptor being associated with a second component capable of providing a detectable signal in response to the binding of a compound to said receptor, with an analytically detectable cytostatin I polypeptide and a compound under conditions to permit binding to the receptor; and

determining whether the compound binds to and inhibits the receptor by detecting the absence of a signal generated from the interaction of the cytostatin I with the receptor.

FIG. 1A

```
       10                            30                            50
CACGAGCTGGAATCTCTCAGCCTCACCTGCCAGACAACACCCCCTCCTTCCTCACCCTGT
       70                            90                       110
TTCCTGCATTCTCCTGAAACCTTCATCCACACAATGCCTCCCAACCTCACTGGCTACTAC
                                   M  P  P  N  L  T  G  Y  Y
      130                           150                       170
CGCTTTGTTTCGCAGAAGAACATGGAGGACTACCTGCAAGCCCTAAACATCAGCTTGGCT
R  F  V  S  Q  K  N  M  E  D  Y  L  Q  A  L  N  I  S  L  A
      190                           210                       230
GTGCGGAAGATCGCGCTGCTGCTGAAGCCGGACAAGGAGATCGAACACCAGGGCAACCAC
V  R  K  I  A  L  L  L  K  P  D  K  E  I  E  H  Q  G  N  H
      250                           270                       290
ATGACGGTGAGGACGCTCAGCACCTTCCGAAACTACACTTTGCAGTTTGATGTGGGAGTG
M  T  V  R  T  L  S  T  F  R  N  Y  T  L  Q  F  D  V  G  V
      310                           330                       350
CAGAAAGGGGAGGTCCCCAACCGGGGCTGGAGACACTGGCTGGAGGGAGAGTTGCTGTAT
Q  K  G  E  V  P  N  R  G  W  R  H  W  L  E  G  E  L  L  Y
      370                           390                       410
CTGGAACTGACTGCAAGGGATGCAGTGTGCGAGCAGGTCTTCAGGAAGGTCAGATAGCCG
L  E  L  T  A  R  D  A  V  C  E  Q  V  F  R  K  V  R
```

MATCH WITH FIG. 1B

MATCH WITH FIG. 1A

FIG. 1B

```
430                                     450                                     470
GAGAGGAGCCAAGATCCCTCCAGACAGCACCAGCTCACAGACGCTCTTGTTGTGCCCCCT
490                                     510                                     530
TCAAGCCCAGATTGTGCCAGGTCAGCTGTCCCTTCCTCTGGCCACCTTTCCTCCCTCTGG
550                                     570                                     590
GTCCCTCCTCACCCCTCCCCGTGTTAATCTGTAACTTGGAGCCCCCAGGACAAAGTCCTT
610                                     630                                     650
TCTCACACTCCACTGCCCAATAGTGACCTCACTTCCAGGTCAAGGTCTGGCGTCCCAAAT
670                                     690                                     710
GAAAGAAGCAGGCAAAGGGAAGGAGCCCCTGAGGACAACCAATCTCCGCTCTCTCCTGTC
730                                     750                                     770
CATTTGACCTCTTCTTTTCCTTCTAAGAAAGAACTAAGCTTTGGGCATTTGGCGATTAGT
790                                     810                                     830
GAAAATTCTATCCTGATGGACTTCTGGAAAACTGTGACTGGGGTTCAACAGTTTAAACAG
850

GGGCTACTGGGGGAAAAAAAA
```

FIG. 2A

```
    M . . . . . . G . . . . . . . . . N . . . Y . . . L . . . . . A .   Consensus #1
    M P D A F L G T W K L V S S E N F D D Y M K A L G V G F A T   Majority
                      10                  20                  30
 1  M P P N L T G Y Y R F V S Q K N M E D Y L Q A L N I S L A V   HTOBH93
 1  M A D A F V G T W K L V D S K N F D D Y M K S L G V G F A T   MDGI
 1  M P V D F T G Y W K M L V N E N F E E Y L R A L D V N V A L   CRBPI
 1  M T R D Q N G T W E M E S N E N F E G Y M K A L D I D F A T   CRBPII
 1  M V D A F L G T W K L V D S K N F D D Y M K S L G V G F A T   FABP
 1  M S N K F L G T W K L V S S E N F D D Y M K A L G V G L A T   Myelin P2

    . . . . . . . . . . . . I . . . G . . . . . . T . S T F . N   Consensus #1
    R K I A S L L K P T K I I E K D G D I I T I K T L S T F K N   Majority
                      40                  50                  60
31  R K I A L L L K P D K E I E H Q G N H M T V R T L S T F R N   HTOBH93
31  R Q V A S M T K P T T I I E K N G D T I T I K T Q S T F K N   MDGI
31  R K I A N L L K P D K E I V Q D G D H M I I R T L S T F R N   CRBPI
31  P K I A V R L T Q T K V I D Q D G D N F K T K T T S T F R N   CRBPII
31  R Q V A S M T K P T T I I E K N G D I L T L K T H S T F K N   FABP
31  R K L G N L A K P T V I I S K K G D I I T I R T E S T F K N   Myelin P2

    . . . . F . . . . . . . . . . . . . . . . . . . . . . . . .   Consensus #1
    T E I S F Q V G V E F D E - - - - T T A D D R K V K S L V T   Majority
                      70                  80                  90
```

MATCH WITH FIG. 2B

FIG. 2B

MATCH WITH FIG. 2A

```
61  Y T L Q F D V - - - - - - - - - - - - - - - - - - - - - -   HTOBH93
61  T E I N F Q L G I E F D E - - - - V T A D D R K V K S L V T   MDGI
61  Y I M D F Q V G K E F E E D L - - T G I D D R K C M T T V S   CRBPI
61  Y D V D F T V G V E F D E - - Y T K S L D N R H V K A L V T   CRBPII
61  T E I S F K L G V E F D E - - - - T T A D D R K V K S I V T   FABP
61  T E I S F K L G Q E F E E - - - - T T A D N R K T K S I V T   Myelin P2

    . . . . . . . . . Q . . . G . . . . . . . . . . . . . . .   Consensus #1
    L D G G K L V H V Q K W - G E E T T R G R E - W L E G G K L   Majority
                      100                 110                120
68  - - - - - - - G V Q K - - G E V P N R G W R H W L E G E L L   HTOBH93
87  L D G G K L I H V Q K W N G Q E T T L T R E - - L V D G K L   MDGI
89  W D G D K L Q C V Q K - - G E K E G R G W T Q W I E G D E L   CRBPI
89  W E G D V L V C V Q K - - G E K E N R G W K Q W I E G D K L   CRBPII
87  L D G G K L V H L Q K W D G Q E T T L V R E - - L I D G K L   FABP
87  L Q R G S L N Q V Q R W D G K E T T I K R K - - L V N G K M   Myelin P2

    . . . . . . . . . V . . . . . . K . . . .                     Consensus #1
    I L E L T H G G V V C T Q V F E K - - A -                     Majority
                      130                 140
89  Y L E L T A R D A V C E Q V F R K V - R                       HTOBH93
115 I L T L T H G S V V S T R T Y E K - E A                       MDGI
117 H L E M R V E G V V C K Q V F K K V - Q                       CRBPI
117 Y L E L T C G D Q V C R Q V F K K - - K                       CRBPII
115 I L T L T H G T A V C T R T Y E K - E A                       FABP
115 V A E C K M K G V V C T R I Y E K - - V                       Myelin P2
```

FIG. 3A

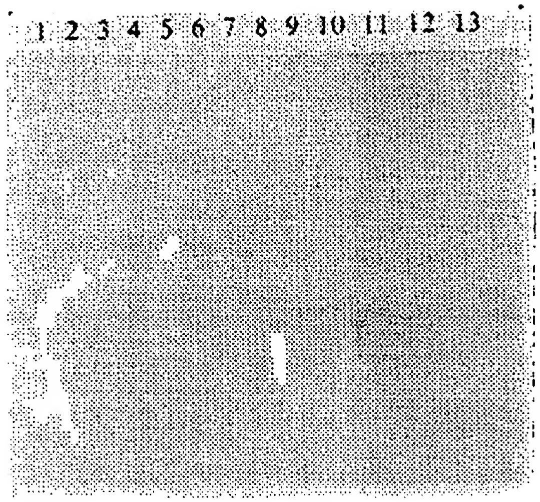

FIG. 3B

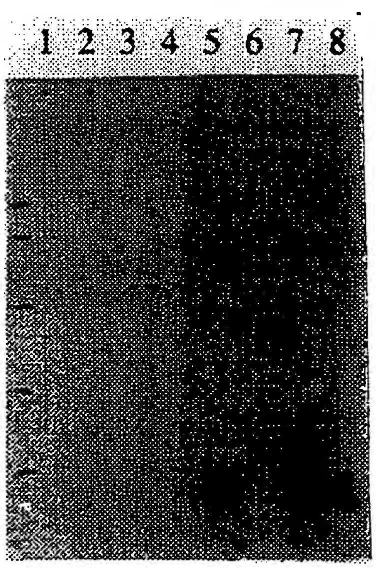

FIG. 3C

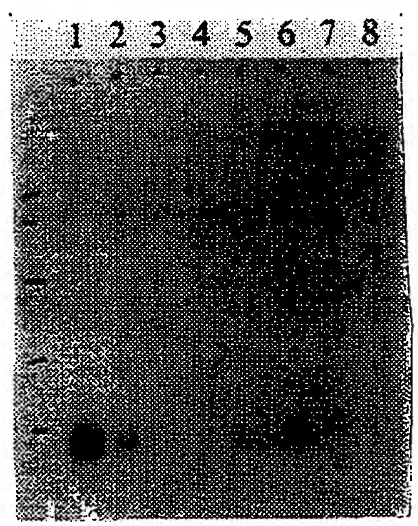

FIG. 4

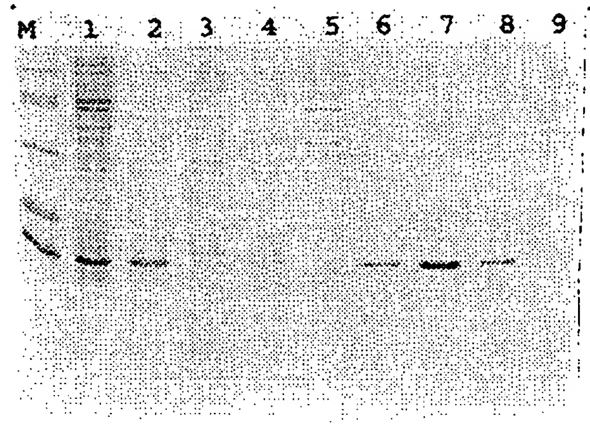

FIG. 6A

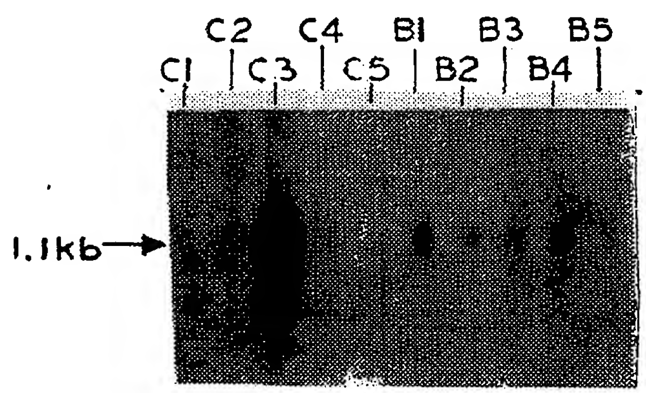

FIG. 6B

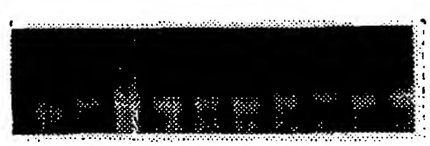

FIG. 5A

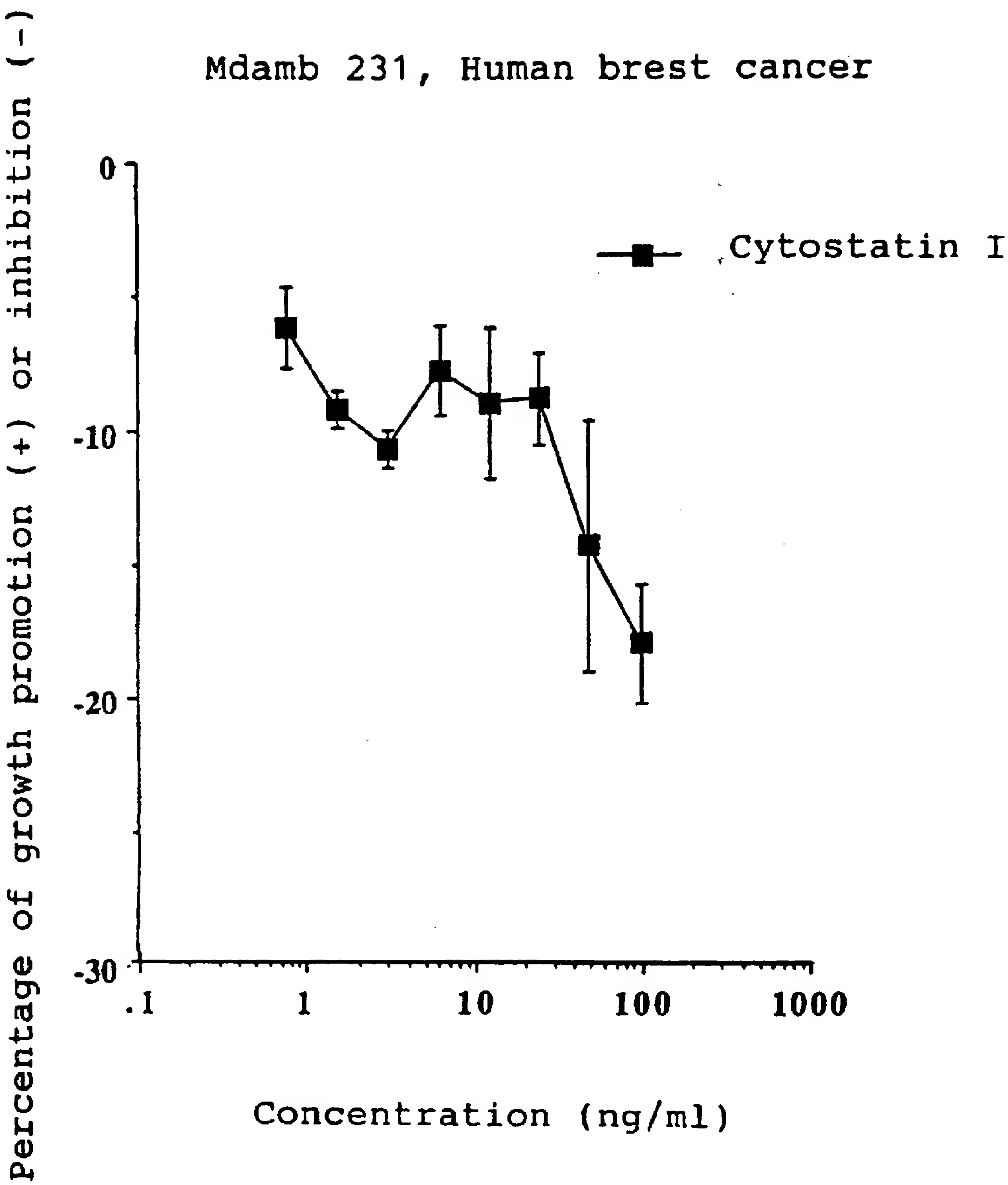
Mdamb 231, Human brest cancer
Percentage of growth promotion (+) or inhibition (-)
0
-10
-20
-30
Cytostatin I
.1
1
10
100
1000
Concentration (ng/ml)

FIG. 5B

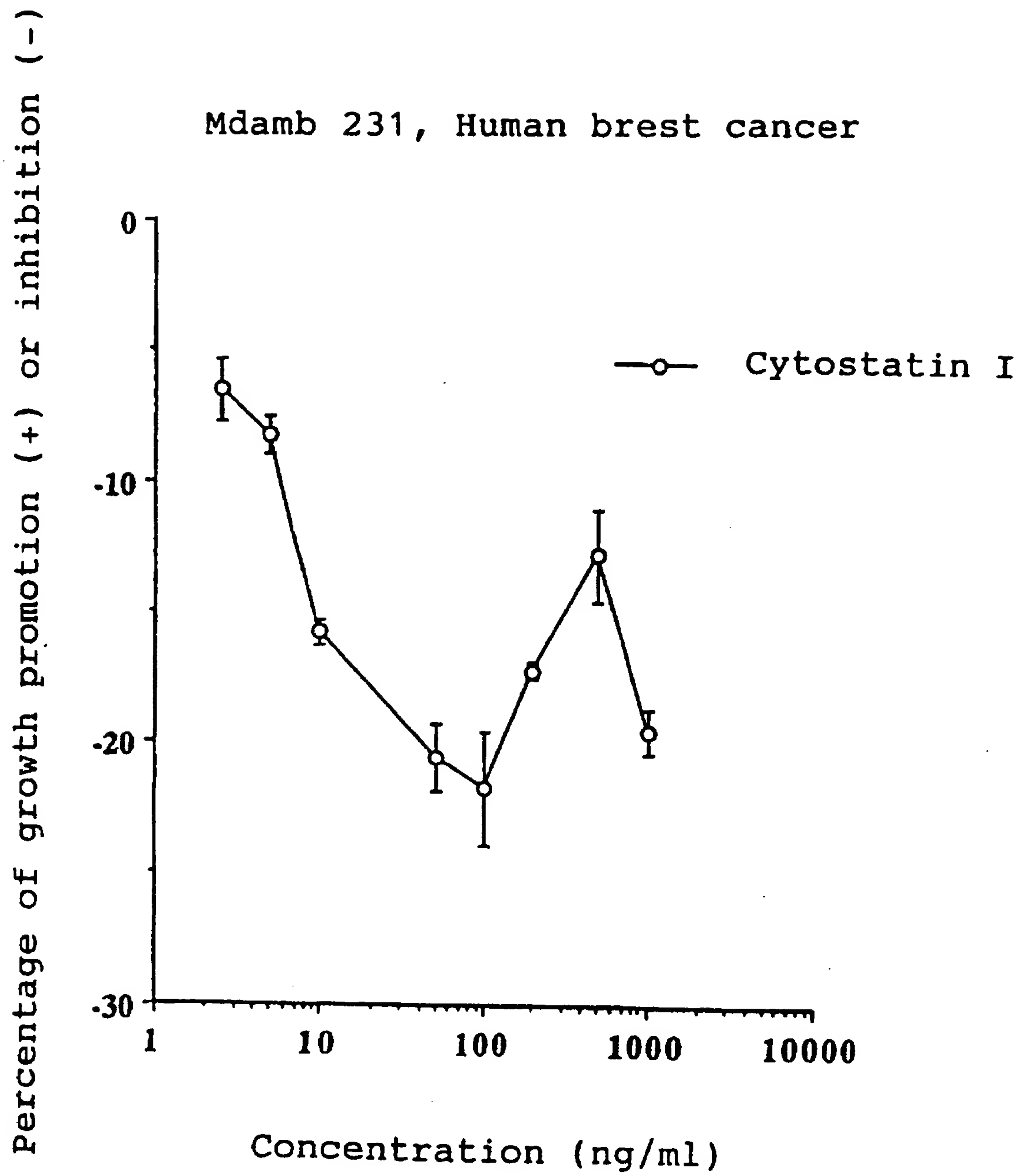
Mdamb 231, Human brest cancer
Percentage of growth promotion (+) or inhibition (-)
0
-10
-20
-30
Cytostatin I
1
10
100
1000
10000
Concentration (ng/ml)

FIG. 5C

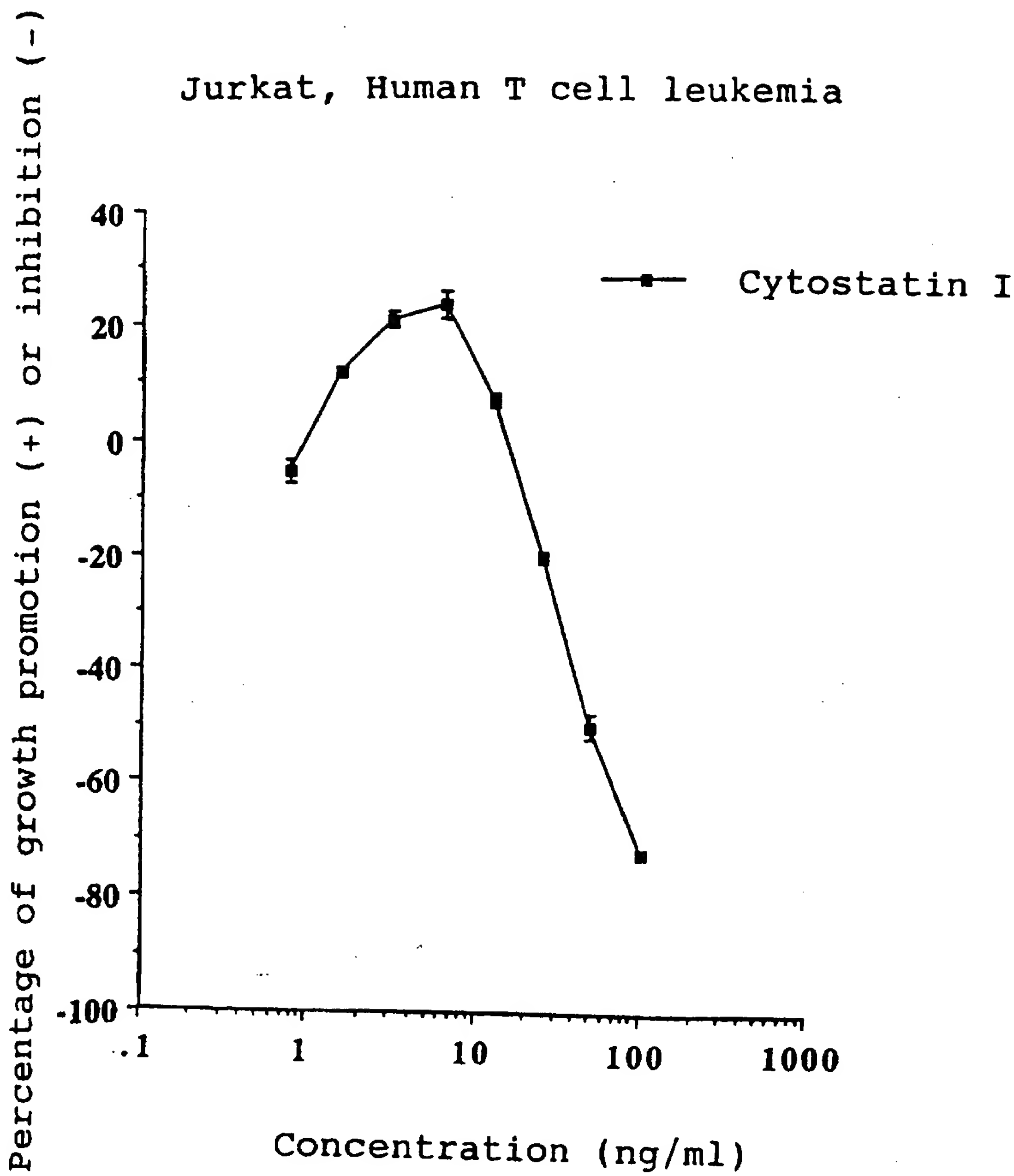
Jurkat, Human T cell leukemia
Percentage of growth promotion (+) or inhibition (-)
40
20
0
-20
-40
-60
-80
-100
Cytostatin I
.1
1
10
100
1000
Concentration (ng/ml)

FIG. 5D

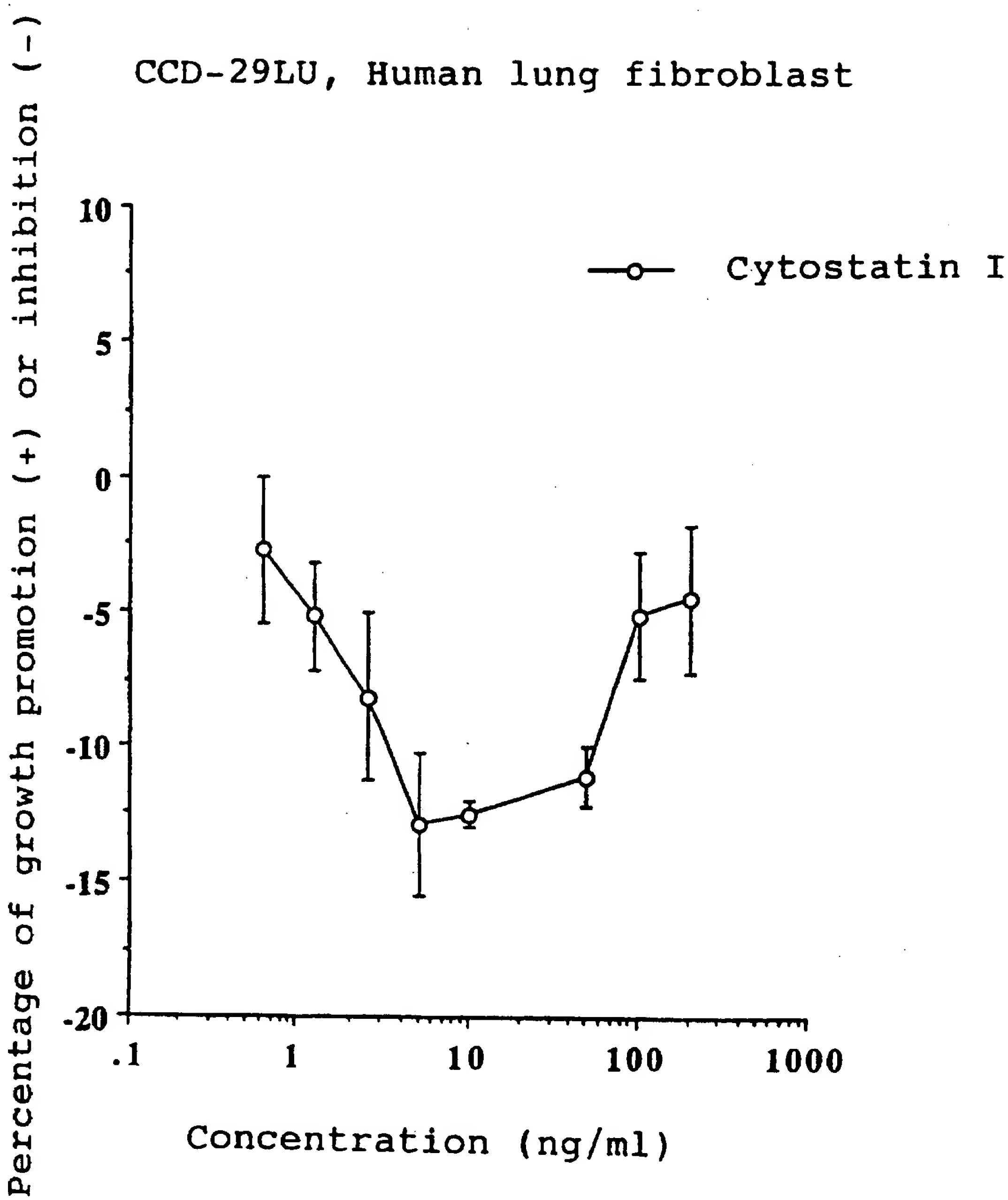
CCD-29LU, Human lung fibroblast
Cytostatin I
Percentage of growth promotion (+) or inhibition (-)
10
5
0
-5
-10
-15
-20
.1
1
10
100
1000
Concentration (ng/ml)

FIG. 5E

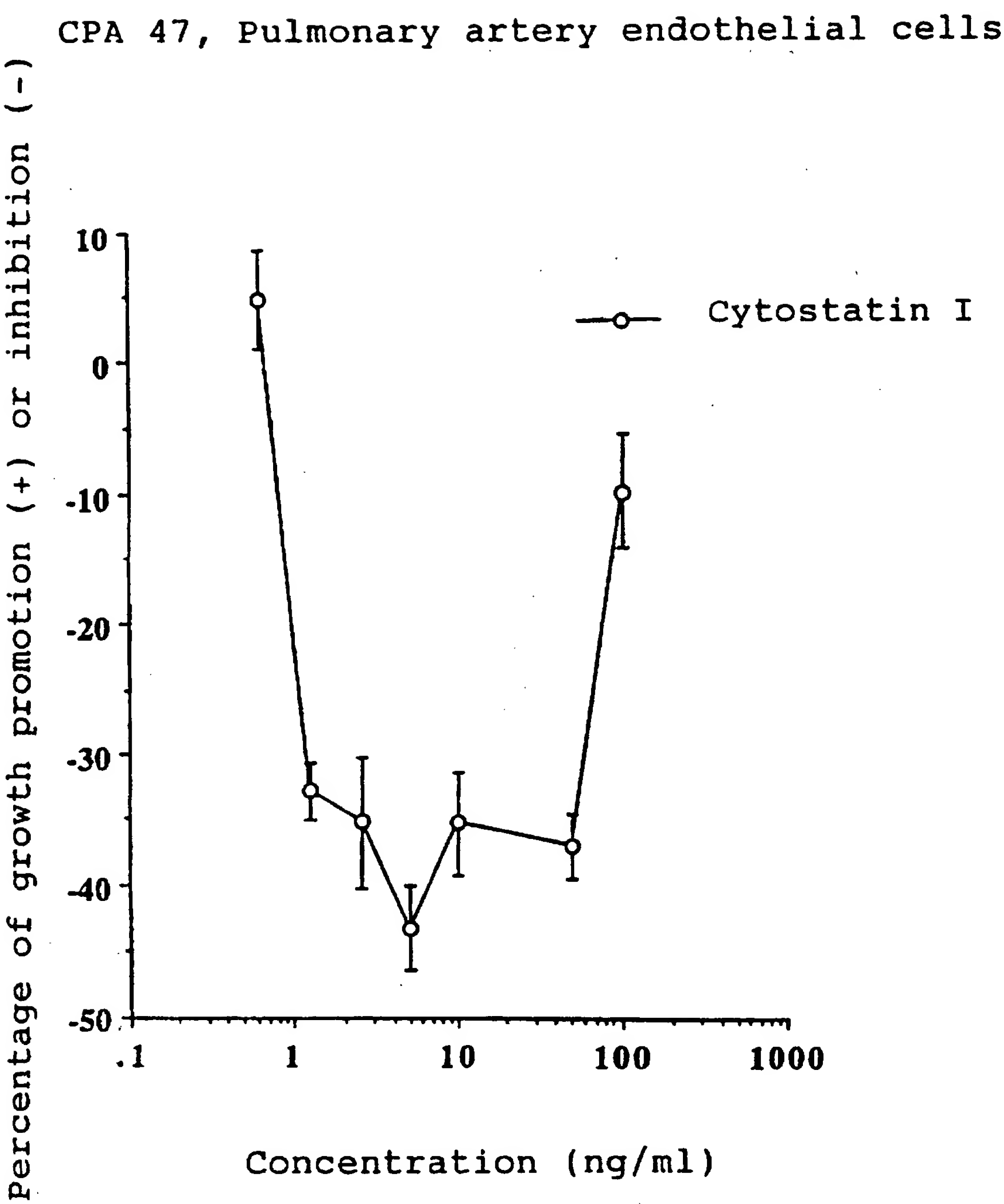
CPA 47, Pulmonary artery endothelial cells
Percentage of growth promotion (+) or inhibition (-)
10
0
-10
-20
-30
-40
-50
.1
1
10
100
1000
Concentration (ng/ml)
Cytostatin I

FIG. 7A

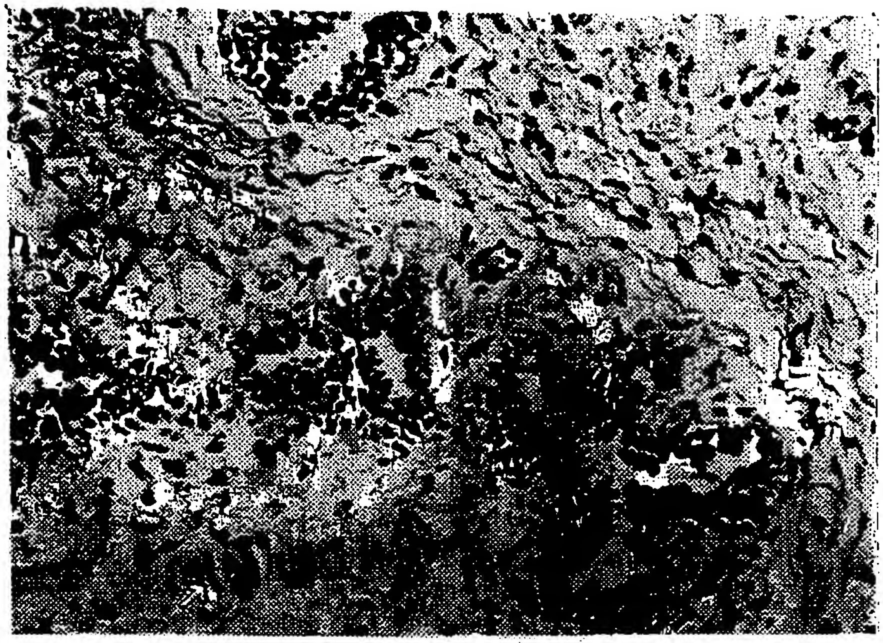

FIG. 7B

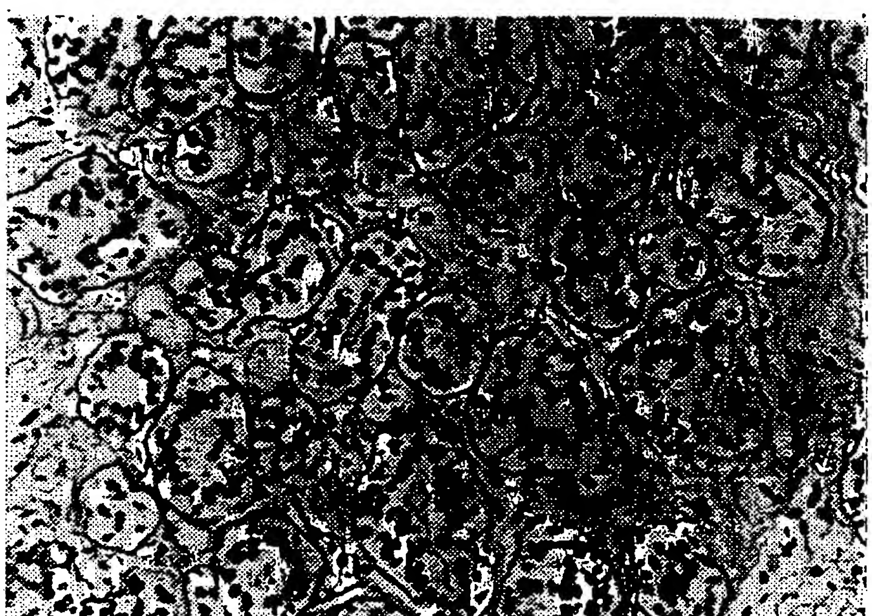

FIG. 7C

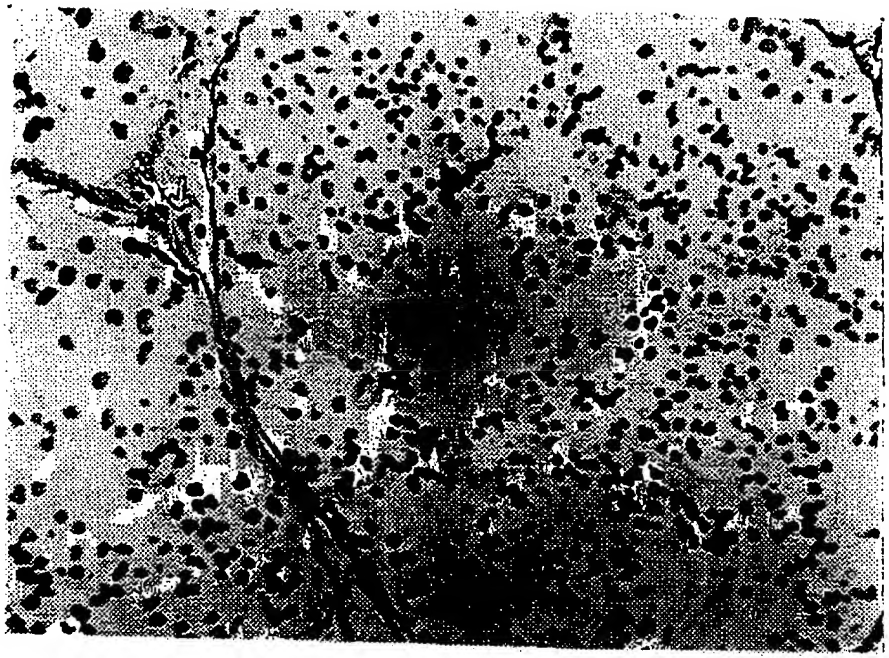

FIG. 7D

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/01640

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :Please See Extra Sheet.
US CL :Please See Extra Sheet.
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 514/2; 530/333, 350, 388.22, 820; 532/23.1, 935/60,66,77

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	The Journal of Antibiotics, Volume 47, No. 5, issued May 1994, Amemiya et al, " Cytostatin, A Novel Inhibitor of Cell Adhesion to Components of Extracellular Matrix Produced by *Streptomyces* sp. MJ654-NF4 I. Taxonomy, Fermentation, Isolation and Biological Activities, pages 536-540.	18 ----- 1-23
A	The Journal of Antibiotics, Volume 47, No. 5, issued May 1994, Amemiya et al, "Cytostatin, A Novel Inhibitor of Cell Adhesion to Components of Extracellular Matrix Produced by *Streptomyces* sp. MK654-NF4 II. Physico-chemical Properties and Structure Determination" pages 541-544, especially see page 542-44.	22

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 MAY 1996	13 JUN 1996

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	JOSEPH CURTIS
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0916

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/01640

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	The Journal of Antibiotics, Volume 48, No.10, issued October 1995, Yamazaki et al, "Screening for Apoptosis Inducers in Microbial Products and Induction of Apoptosis by Cytostatin", pages 1138-40, especially pages 1138-1139.	23

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/01640

A. CLASSIFICATION OF SUBJECT MATTER:
IPC (6):

A01N 37/18, 43/04; A61K 38/00; C07H 21/02; C07K 1/00, 5/00, 16/00; C07G 17/00; C12N 1/20, 15/00; C12P 21/06; C12Q 1/68; G01N 33/53; H01R 13/62

A. CLASSIFICATION OF SUBJECT MATTER:
US CL :

435/6, 7.1, 7.2, 69.1, 240.2, 252.3, 320.1; 514/2, 44; 530/333, 350, 388.22, 820; 536/23.1, 935/60,66,77; 536/23.1

B. FIELDS SEARCHED
Electronic data bases consulted (Name of data base and where practicable terms used):

APS, BIOSIS, CAPLUS, EMBASE, PHAR, USPATFIL, WPIDS, JPIO
search terms: cytostatin, cytostatin I, disease, vector, plasmid, screen, diagnosis, inhibition, polypeptides, protein, gene, expression, underexpression, bind, human, and cell.